Teacher Perception of Locally Developed Value-Added Measures

A Dissertation Submitted to the Graduate School in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

By

Jack S. Pettit

Dr. Serena Salloum – Advisor

Ball State University

Muncie, Indiana

December 2019

Approval Page

ABSTRACT

DISSERTATION: Teacher Perception of Locally Developed Value-Added Measures

STUDENT: Jack S. Pettit

DEGREE: Doctor of Education in Leadership

COLLEGE: Teachers College

DATE: December 2019

PAGES: 117

Value Added Measures (VAMs) have been developed to measure teacher, school, and district effectiveness for state and federal accountability purposes. Previous VAM research has focused on defining, justifying, critiquing, and exploring the statistical validity of VAMs. This study will attempt to bridge the literature gap regarding implications of VAMs on teaching practice. The research methods utilized in the study are mixed, and a case study format is used. The population of the study consists of teachers in the United States who work in a school in which a VAM is utilized. However, the case study is delimited to a single setting. The setting for the study is a high-achieving, suburban, high school in Northwest Indiana with approximately 2,750 students. The quantitative research question of the study examined how the adoption of a single high school's VAM shaped and influenced teacher practice. Quantitative results indicate that teacher practices only changed in two areas (1) the use of strategies in which teachers asked students to revise knowledge and, (2) teachers developing lessons in which students had to identify their own learning resources. Qualitative research examined any changes that occurred since VAM implementation. Teacher responses indicated that there have been changes in teacher practice, but not due to VAM. Technology and student performance on high-stakes, end-

of-course assessments were cited as the two factors most responsible for change in teacher practice.

DEDICATION

To my dad: Jack Pettit, Sr. (deceased), who would have loved to have been here to see this project get completed.

ACKNOWLEDGEMENTS

To my mentor and committee chair, Serena Salloum, thank you. I could not have asked for a more knowledgeable and supportive mentor throughout my graduate coursework and writing the dissertation. I will never be able to repay you for the thoughtful feedback you have given me throughout this process. You are everything that a teacher should be and I find myself modeling my work after you.

To my supportive, and dedicated committee members, Marilyn Quick, Kendra Lowery, and Jill Bradley-Levine, thank you for pushing me to become a better researcher and scholar. This project has improved with each bit of advice you have shared along the way.

Thank you to the BSU faculty that have provided incredible insight into the field of Educational Leadership. I hope to continue to rely on your friendship and expertise in the years to come. To my advisor, Marilyn Quick, you are an amazing woman. Your positive affect that blends encouragement with a commitment to high standards has served me and countless others going through this program extremely well. To John Ellis, it has been a treat to gain insight into your knowledge on finance, educational legislation, and the business side of K-12 education. To Lynn Lehman, I am lucky to have had the chance to take your course and I value the mentoring you have provided. To Lori Boyland and Jeff Swenson, your positive outlook and encouragement early on in my coursework was appreciated. I would have loved working in a school or district under your leadership. You have all been tremendous role models, thank you.

To my professional mentors, Michael Hazen, Bill Dorulla, Dr. Steve Sprunger, and Dr. E. Ric Frataccia (deceased), your support and encouragement throughout the years has meant a lot to me; thank you.

To Dr. Teresa Eineman, my superintendent and mentor for the past twelve years, thank you for supporting me on both a personal and professional level. This project would not have come to fruition without you believing in this work. The professional culture that you have created in the Crown Point Schools in which teachers and administrators engage in a continuous cycle of improvement is one in which I have been proud to be a part of.

To those in the Crown Point schools that have engaged in this work as thought partners and teammates, Dr. Eric Ban, Jim Hardman, Deb Cuffia, Mark Gianfermi, Deb Ciochina, Russ Marcinek, Robert McDermott, Jeremy Walker, Sabrina Ramirez, Annie Lorek, John Lambersie, Rachele Raloff, Dan Hadary, Jacob Adams, Jerome Flewelling, Jim Ingelhart, and Josh Graegin. Thank you.

To Ryan Collier, your work with ACT and the development of the growth model was inspiring. I have enjoyed our many conversations through the years about how to position this work with faculty. Your many talents and roles, both formal and informal at CPHS (AP Latin, AP Research, Pre-AP Pre-Calculus teacher, and teacher association leadership), have played a large role in creating a positive professional culture within our school.

To Jeff Allen and ACT, your willingness to pursue initiatives in the best interests of teaching and learning is admirable. ACT provided access to national data sets that contributed to the development of the growth model without asking for anything in return. Thank you.

And last, but certainly, not least, thank you to my family. To Karin, for your love and support during this project. I know it always seems like one more thing, coaching, graduate school, etc. Thank you for always being supportive. To my mom, Pam, and Bill and Joan, thank you for the encouragement! To my children, Will and Avery, no more long drives to Indianapolis or Muncie, but the pursuit of learning more never ends.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1

INTRODUCTION

Teacher Perception of Locally Developed Value-Added Measures

The recent history of American education is dominated by the concept of accountability (Achinstein & Ogawa, 2012; Atkinson, 2015; Busuioc & Lodge, 2016; Cooley Fruehwirth, 2013; Hartney & Flavin, 2014; Lee & Reeves, 2012; Shaw, 2016). Increasingly, the accountability of the American school system has been defined in terms of the results of high-stakes testing (Block, 2015; Coburn, Hill, & Spillane, 2016; Glover, Reddy, Kettler, Kurz, & Lekwa, 2016; Lewis & Hardy, 2015; Plank & Condliffe, 2013; Singh, Märtsin, & Glasswell, 2015; Witte, Wolf, Cowen, Carlson, & Fleming, 2014). The underlying performance-based aspect of the accountability era in American education is not necessarily new; as early as the 1950s, as a response to the general panic created by the Soviet ability to put human beings into orbit before the United States could do so, the American education system began to redefine itself on the principle of improving student performance (Shapiro, 2015). However, from the 1990s onwards, a legislative and social consensus has emerged that schools (and, in particular, teachers) are accountable for student performance (Block, 2015; Coburn et al., 2016; Glover et al., 2016; Lewis & Hardy, 2015; Plank & Condliffe, 2013; Singh et al., 2015; Witte et al., 2014). In the decades of educational reform that followed the crisis engendered by the Sputnik launch, American policy-makers and educators emphasized a more holistic approach to accountability, one in which individual students, parents, teachers, principals, schools, school systems, and governmental structures were all accountable for the improvement of student performance (Shapiro, 2015).

However, beginning in the 1990s, the onus of accountability began to shift more towards schools, and, in particular, teachers (Block, 2015; Coburn et al., 2016; Glover et al., 2016; Lewis & Hardy, 2015; Plank & Condliffe, 2013; Singh et al., 2015; Witte et al., 2014). In 2001, the passage of the No Child Left Behind (NCLB) Act officially marked the beginning of the current accountability era in American education (Chakrabarti, 2014; Coburn et al., 2016; Grissom, Nicholson-Crotty, & Harrington, 2014; Ladd, 2017; Lavery, 2016; Menken & Solorza, 2014; Reback, Rockoff, & Schwartz, 2014). Among its other provisions, NCLB mandated an annual test of academic progress based on the administration of a standardized test and tied the funding and other administrative aspects of schools to meeting the so-called Adequate Yearly Progress (AYP) metric of standardized test score achievement (measured in terms of the percentage of students able to meet or exceed performance norms in reading and mathematics) (Chakrabarti, 2014; Coburn et al., 2016; Grissom et al., 2014; Ladd, 2017; Lavery, 2016; Menken & Solorza, 2014; Reback et al., 2014).

Value-added measures (VAMs) emerged in the aftermath of NCLB as a means of quantifying the contribution of teachers to student achievement (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber, Goldschmidt, & Tseng, 2013; Henry, Rose, & Lauen, 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher, Selig, Cosbey, & Thorstensen, 2014). Because NCLB defined AYP as the main metric of American educational progress, the responsibility and accountability of the teacher were increasingly redefined in terms of what teachers contributed to the academic performance of their students, and this redefinition took place through the VAM paradigm (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014;

Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014).

The common phrase 'teaching to the test' captures the potential impact of VAMs on the practice of both teachers and administrators (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014). After NCLB, teachers' missions were increasingly redefined in terms of the importance of tests, and in many schools and districts, the collection and analysis of VAM data was adopted as a means of determining whether teachers were, in fact, making significant contributions to the performance of their students (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014). Subsequent VAM research and practice has taken many dimensions; in particular, researchers have focused on defining, justifying, critiquing, and exploring the statistical validity of VAMs (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014). However, what is not generally known is whether the imposition of VAMs, as part of a generalized school and school district response to high-stakes testing, has resulted in a genuine transformation in teachers' practices.

In terms of programmatic logic, the rationale for VAMs, as for NCLB in general, was to provide an incentive structure for teachers and principals, among other personnel, to change their practices (Block, 2015; Chakrabarti, 2014; Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Glover et al., 2016; Goldhaber et al., 2013; Grissom et al., 2014; Henry et al., 2014; Ladd, 2017; Lavery, 2016; Menken & Solorza, 2014; Paufler & Amrein-Beardsley, 2014; Plank

& Condliffe, 2013; Price, 2014; Reback et al., 2014; Saultz & Saultz, 2017; Singh et al., 2015; Steinbrecher et al., 2014). One longstanding critique of the American educational system has been that neither teachers nor administrators have been sufficiently hardworking, attentive to student needs, or otherwise equipped to maximize the academic results of students (Hampden-Thompson, Guzman, & Lippman, 2013; Han, 2016; Hitt & Tucker, 2016; Peurach & Marx, 2010; Theron, 2013). This critique existed long before NCLB; it can be traced well into the nineteenth century (Hampden-Thompson et al., 2013; Han, 2016; Hitt & Tucker, 2016; Peurach & Marx, 2010; Theron, 2013). In the post-NCLB era, however, the critique of teachers' and administrators' practices was tied, at least in theory, to measurable performance (Block, 2015; Chakrabarti, 2014; Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Glover et al., 2016; Goldhaber et al., 2013; Grissom et al., 2014; Henry et al., 2014; Ladd, 2017; Lavery, 2016; Menken & Solorza, 2014; Paufler & Amrein-Beardsley, 2014; Plank & Condliffe, 2013; Price, 2014; Reback et al., 2014; Saultz & Saultz, 2017; Singh et al., 2015; Steinbrecher et al., 2014). In this context, the NCLB can itself be understood as a means of creating an incentive structure to alter the practices of teachers and administrators. Scholars have paid considerable attention to the question of whether VAMs and NCLB focus on accountability have improved students' academic outcomes (Block, 2015; Chakrabarti, 2014; Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Glover et al., 2016; Goldhaber et al., 2013; Grissom et al., 2014; Henry et al., 2014; Ladd, 2017; Lavery, 2016; Menken & Solorza, 2014; Paufler & Amrein-Beardsley, 2014; Plank & Condliffe, 2013; Price, 2014; Reback et al., 2014; Saultz & Saultz, 2017; Singh et al., 2015; Steinbrecher et al., 2014). Attention has also been paid to how accountability policies, including NCLB, have influenced local district and school practices (Dee, Jacob, & Schwartz, 2013). The assumption is that local measures are developed,

implemented, and monitored to increase the likelihood of student achievement and growth gains on standardized tests.

The logic of NCLB theorized that changing incentive structures would result in an alteration in the practices of teachers (among other stakeholders), and that these changes would facilitate an improvement in student performance (Amrein-Beardsley & Holloway, 2019; Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014). Thus, the relationship between NCLB (and its specific mechanisms of personnel measurement, especially VAMs) and the desired improvement in student performance cannot be validly measured without first studying the relationship between VAMs and the practices of teachers.

Statement of the Problem

The problem addressed in this dissertation has local as well as national components, but at both the local and national levels, the problem can be conceptualized in the same manner. For high-stakes testing to be utilized in the manner intended by NCLB and subsequent legislation, policy changes have to create an incentive for teachers to alter their practices (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014). If teachers do not change their practices, there is no plausible mechanism whereby they can influence student achievement. As the research literature indicates, the actual practices of educational personnel, not their emotions or other intangible factors, are associated with improved student performance (Barge & Loges, 2003; Borg, Mary, & Harriet, 2012; Davis, Gabelman, & Wingfield, 2011; Ford & Sassi, 2014; Gambrell, 2015; Griner & Stewart, 2013;

Hill, Rowan, & Ball, 2005; Koh, Steers, & Terborg, 1995; Milner & Howard, 2004; Mojavezi & Tamiz, 2012; Shevalier & McKenzie, 2012; Wenglinsky, 2002; Xu, Coats, & Davidson, 2012).

In this context, the problem addressed in the study is that there is insufficient empirical knowledge about the link between VAMs and changes in the practices of teachers. In the absence of this knowledge, legislators and policy-makers cannot be certain that statutes such as NCLB, and now ESSA (2015), are having their intended effect of creating an accountability-based educational system in the United States. In the local context, a school that has created a VAM structure does not know whether this structure has been effective in changing the practices of teachers. The absence of local knowledge means that decision-makers lack the empirical support necessary to determine whether to keep the VAM in place, discard the VAM, or change the VAM. If, for example, the VAM has not resulted in a meaningful change in teacher practice, then local decision-makers would have a rationale to change the VAM or abandon it. On the other hand, if the VAM resulted in a change in teacher practice, and if this change were held to be positive, then decision-makers would have an empirical basis for retaining the VAM.

Purpose of the Study

The purpose of this mixed-methods case study was to examine how the adoption of a single high school's VAM shaped and influenced teacher practice. The presumption of the study was that the imposition of the VAM was likely to create changes in the practices of teachers. The purpose of the study was be achieved by answering two questions: (1) What changes in teacher practices have occurred since the administration of a VAM in a particular school? (2) Did VAM alter the practices of teachers? Why or why not?

Nature of the Study

The study was a mixed-methods, case study, post-positivistic and pragmatic in nature. The study was quantitative insofar as it was based on the mathematical measurement of the relationship between a dichotomous predictor variable (that of the imposition of a VAM) and subsequent teacher practices. The study was qualitative insofar as it relied on the examination of personal narratives to answer context-dependent questions about the rationales for teacher and administrator response to a VAM imposition. The study was post-positivistic in nature insofar as the utilization of mathematical models was not taken to constitute a presumption about the actual nature of educational personnel's practices. The study was pragmatic in nature as the analysis and interpretation of the qualitative data were used to inform whether to continue or adjust VAM. Finally, the case study was situated in an affluent suburban context that is not necessarily reflective of other settings.

Theoretical Framework

According to Henderikus, a theory "is normally aimed at providing explanatory leverage on a problem, describing innovative features of a phenomenon or providing predictive utility" (Henderikus, 2010, p. 1498). These criteria ought to be satisfied by any theoretical framework applied to the question of how VAMs impact, or fail to impact, teacher practices. An acceptable theoretical framework should provide (a) descriptions of a phenomenon or phenomena, (b) explanations of a phenomenon or phenomena, and (c) predictions about the possible future state of a phenomenon or phenomena in relation to some change (Henderikus, 2010). The theoretical framework chosen for this study was Ryan and Deci's (2000) self-determination theory. Ryan and Deci wrote that self-motivation theory was based on "the postulate of three innate psychological needs—competence, autonomy, and relatedness—which when satisfied yield

enhanced self-motivation and mental health and when thwarted lead to diminished motivation and well-being" (Ryan & Deci, 2000, p. 68). Thus, according to Ryan and Deci, self-determination is the combined expression of free decision-making (autonomy), the exercise of independent skill (competence), and the feeling of being cared for and valued by others (relatedness).

VAMs can be conceptually explored through each of the components of self-determination theory. Insofar as VAMs might be perceived by teachers as an attempt to control the process and content of teaching, and thereby the freedom of the individual teacher, VAMs might reduce autonomy. The drive towards teacher measurement has been closely aligned to the need, as expressed in educational policy, to prepare children for higher levels of achievement on various tests, whether the federal tests mandated by NCLB or the state tests in the era of the Every Child Succeeds Act (ESSA) of 2016. To the extent that preparing children for success in high-stakes testing requires a standardization of teaching, and to the extent that the standardization of teaching reduces the freedom of the individual teacher, self-determination theory predicted that the utilization of VAMs would harm self-determination by reducing teacher autonomy.

In self-determination theory, autonomy, competence, and relatedness are tied to the motivation and ability of an individual to engage in a complex task such as teaching (Ryan & Deci, 2000). Therefore, damage to teacher autonomy, or other components of self-determination, would likely result in a decrease in both motivation and ability to perform, both of which could, in turn, negatively alter teacher practices. In terms of competence, it is possible that the adoption of a VAM would signal to teachers that their competence will be measured through a single metric, that of student performance, which might not be under the sole, or even

the predominant, control of the teacher. To the extent that VAMs reduce the various manifestations of teacher competence into a single, possible unreliable or invalid measurement, self-determination predicts that the utilization of VAMs would demotivate teachers by signaling that non-VAM-based forms of teacher competence are not important.

Finally, self-determination theory can be applied to VAMs in terms of relatedness. The relatedness component of self-determination reflects the individual need to feel cared for and valued by others (Ryan & Deci, 2000). To the extent that VAMs might make teachers feel that they will be rigidly judged and subjected to close control, VAMs might reduce the feeling of teacher relatedness and thereby do additional damage to teachers' motivation and ability. The punitive nature of VAMs may also cause teachers to band together in professional learning communities to cope with the pressures associated with high-stakes testing in this era of accountability.

Research Questions and Hypotheses

The quantitative research question and accompanying hypotheses of the study were as follows:

RQ1: What changes in teacher practices have occurred since the administration of a VAM in a particular school?

HI_0: Teacher reports of their practices before the administration of a VAM in a particular school were not statistically different to teacher practices after the administration of a VAM in a particular school.

HI_A: Teacher reports of their practices before the administration of a VAM in a particular school were statistically different to teacher practices after the administration of a VAM in a particular school.

The qualitative research questions of the study are as follows:

RQ2: Did VAM alter the practices of teachers? Why or why not?

Definition of Terms

The main technical term in this study that required definition is that of a value-added measure, or VAM, but teacher practice required definition as well.

Teacher practice. For the purposes of this study, teacher practice was defined in terms of teachers' orientations measured on the Approaches to Teaching Inventory (ATI-R) (Trigwell, Prosser, & Waterhouse, 1999), which has been described in greater detail in the third chapter of the study. The ATI-R has two formal components, student-focused and self-focused practices, but, according to a previous study (Goh, Wong, & Hamzah, 2014) that utilized the ATI-R, the ATI-R has as many as five sub-constructs that emerge through exploratory factor analysis. This aspect of teacher practice is discussed at greater length in the third chapter of the study.

VAM. A VAM was defined as an evaluation standard that purports to calculate an individual teacher's contribution to the academic success of a student (Gansle et al., 2015, p. 107).

Significance of the Study

The study's significance can be understood in terms of local practice as well as in general scholarly terms. One of the main gaps in the literature on VAMs (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014) was the absence of empirical analyses on the relationship between VAM adoption and the changes, if any, in the practices of educational personnel. Insofar as the NCLB and accompanying legislation was designed to alter education personnel's practices through the existence of motivation structures

of VAMs, this gap is highly important. Without knowing whether VAMs were associated with changes in the practices of teachers, the overall success of NCLB and its associated motivation structures could not be appropriately measured. In terms of the policy evaluation of NCLB, more empirical research on the link between VAMs (or equivalent motivational structures) and the dependent variable of educational personnel's behavior was necessary.

The significance of the study can also be understood in terms of the local context of the study. The school that was the research setting for the study had not yet learned whether its customized VAM succeeded in altering the practices of teachers. In the absence of this knowledge, the school lacked the information necessary to decide whether it ought to retain, discard, or change the VAM. By providing empirical insight into the relationship between the school's VAM and the changes, if any, in the practices of teachers, the study provided decision makers in the school and the district with the kinds of information necessary to make evidence-based decisions about the VAMs.

Structure of the Study

The purpose of the first chapter of the study is to introduce the research problem and its general context. The remainder of the study is structured as follows. The second chapter, the review of literature, contains an extensive discussion of VAMs and their role in American education. The third chapter, the methods, consists of a description and defense of the pertinent components of the research methodology and design of the study. The fourth chapter consists of the empirical findings of the study, and the fifth chapter comprises the conclusion of the study.

Positionality Statement

In the 12 years prior to collecting data for the study, I worked as an administrator in the same high school. For the first 5 years, I served as an assistant principal. At the conclusion of the study (July 2019), I had just finished my 7th year as principal. During this time, I led school improvement initiatives that emphasized student skill development in core areas as measured on standardized exams for student, teacher, and school accountability. Many of these initiatives were developed in response to legislation such as: NCLB, PL-221, and ESSA. As a result, I had become well versed in aligning curriculum and standards to end-of-course assessments.

During this time period, end-of-course assessments required of high school students in Indiana were positioned as graduation qualifying exams. Students needed to earn a minimum score at the end of their 10th grade year in both English and mathematics as a condition for graduation. Between 75-85% of students at my school earned a minimum score on the graduation qualifying assessment on their first attempt and 97-98% of students graduated from high school. However, the minimum competency component required for graduation did not resonate with students and parents. As a result, there was a perceived lack of relevance in the assessment. In 2008, the school implemented a longitudinal testing system (ACT and ACT PLAN) to help track individual student achievement and growth as a way to better inform students and parents on academic progress. The alignment to a college entrance exam allowed both teachers and administrators to have meaningful conversations with students and parents on current levels of academic achievement, growth trajectories, and strategies to improve upon a student's projected path. Soon after adopting the longitudinal assessment program, administration started having conversations with ACT personnel about growth modeling (VAM) strategies that were emerging nationwide. Of particular interest, was the development of a VAM

to be used for school improvement purposes. Over the next six years, ACT partnered with the school to develop a VAM for the following purposes: compare student value-added to that of like-demographic schools (measure of school effectiveness), calculate annual value-added by course and teacher; measuring effectiveness of interventions (institutional research), and compare value-added across cohorts to that of like-demographic schools (measure of school improvement). The VAM was implemented during the 2014-15 school year. ACT longitudinal exams and VAM calculations were still being utilized at the conclusion of the study for the aforementioned reasons. VAM introduction at the school was not consistent with how VAM measures had been introduced throughout the country (for accountability purposes). I believe this is an important distinction. Anecdotally, I observed that the school had a professional culture in which teachers were empowered to engage in professional dialogue without the threat of negative repercussions or punitive action. There was a high-level of trust between teachers and administrators. As a result, teachers were regularly engaged in data analysis and decision making relative to school improvement goals.

I do not believe that one-time test scores should be used as the sole indicator for student, school, and teacher performance. I do believe that student test performance should be used to inform school improvement initiatives as schools leverage available resources to promote student achievement and growth. The VAM implemented in the school in which I served was designed for school improvement purposes. As a result, I believe that teachers responded to survey and interview questions differently than if the VAM were used to inform their evaluation. Even so, the power dynamic that existed between me (principal) and the teachers in the school (research participants) cannot be ignored. I believe it is possible that teachers in the school may have thought that I was a proponent of VAMs, without knowing the extent, and answered questions on

both the survey and follow-up interview in such a way that may have been looking to gain my approval. I developed the informed consent with this in mind and attempted to remove any bias that may or may not have existed amongst the teachers in the school. My hope for the study was to further school improvement initiatives within the school and also to expand upon the body of research on use of VAMs.

CHAPTER 2

REVIEW OF LITERATURE

The purpose of this mixed-methods case study was to examine how the adoption of a single high school's VAM shaped and influenced teacher practice. The main objective of the review of literature was to bring theoretical and empirical perspectives to bear on the focus of the study and to identify gaps in the literature that can justify additional analysis. The literature review is structured as follows. First, the purposes of the study and research questions are reiterated. Second, the history of VAMs in the United States is presented. Third, the theoretical framework of the study is discussed. Fourth, arguments for VAMs are presented. Fifth, arguments against VAMs are presented. Sixth, there is a review of evidence related to changes in teacher practice as a function of changes in assessment. Seventh, points of practitioner and scholarly consensus on the current state and future trajectory of VAMs are identified. Eighth, gaps in the literature are identified.

History of Value-Added Models in America

The notion that school effectiveness can be measured by evaluating student outputs can be traced to the Coleman Report (1966). The study, commissioned by the United States Department of Education, was the first of its kind in concluding that variations in school quality showed little association when measuring inputs (per pupil expenditures, size of library, etc.) to determine levels of educational attainment. Furthermore, differences in students' family backgrounds, by comparison, showed a substantial association with achievement (Coleman, 1966). At the time of the report's release, American's public schools were recovering from misguided reform efforts (large-scale national curriculum, open plan schools, and individualized instruction) and a period of stagnation followed that lasted throughout the 1970s (Fullan, 1993).

By the early 1980s, society had become frustrated with the state of public education (Bracey, 2002). *A Nation at Risk*, published in 1983, confirmed the "widespread public perception that something [was] seriously remiss in our educational system" (Bracey, 2002, p.1). The report's Cold War rhetoric produced a Sputnik-type occurrence that the public, and in turn policy makers, could point to a scapegoat to blame for America's economic troubles. While *A Nation at Risk* did not address any Reagan-era education agenda items, such as vouchers, tuition tax credits, restoring school prayer, and abolishing the U.S. Department of Education, it had plenty to serve as a catalyst for Reagan's re-election in 1984 (Bracey, 2002). The irony was that the report called for nothing new, only more of the same: more science, more mathematics, more computer science, more foreign language, more homework, more rigorous courses, more time-on-task, more hours in the school day, more days in the school year, more training for teachers, and more money for teachers (Bracey, 2002). Despite the dire predictions of national economic collapse without educational reform, America's economy soared due to contributions from students (today's workforce) in public schools at the time the report was published. This suggested that nothing was wrong with the educational system. The attempt to eliminate the Department of Education was unsuccessful. But, a crisis (*Nation at Risk*) was 'developed' to further the Republican political agenda (Bracey, 2002; Riley & Peterson, 2008).

State departments of education responded to *A Nation at Risk* by defining grade-level content standards and developing standardized exams to measure student performance (Bishop & Mane, 2001). Theoretically, standardized exam performance assists students, parents, and educators to monitor progress on mastery of learning targets at various points in a student's schooling (Archbald & Newmann, 1988). Most states took this a step further by requiring students to pass a minimum competency exam as a condition for high school graduation (Bishop

& Mane, 2001). Society appeared to have found what it was looking for: a way to determine educational effectiveness by examining outputs (student performance). Predictably, it did not take long for policy makers to turn to standardized exams to quantify student, teacher, and school performance.

What Coleman foreshadowed in the mid-1960s came to fruition in the 1980s with *A Nation at Risk* serving as the catalyst. Researchers started to develop different ways to measure student performance and its relationship on school and teacher effectiveness. In the mid-1980s, William Sanders developed one of the first statistical models in which school and teacher effectiveness could be quantified by measuring student growth over time (Sanders & Horn, 1994). The difference in year-to-year student performance on standardized exams became the "gain" score for each student. Predicted student performance levels were established to determine whether or not a student, teacher, or school met year-to-year expectations.

Sanders and Horn (1994) rationalized that background factors (e.g. educational attainment of parents, family income levels, individual student ability, race) known to impact student achievement were buried in the initial test score. Sanders' work was legitimized in 1991, when the Tennessee state legislature passed the Tennessee Educational Improvement Act, which mandated that school effectiveness would be measured using results from a mixed-model value-add growth system. Sanders and Horn (1994) defended the development of this school accountability model using the same rationale first documented in the Coleman Report. This was to be done by measuring the product (outputs measured by student performance) of the educational experience rather than trying to evaluate schools and teachers through the examination of the process (inputs evaluated by an external team of experts).

There were numerous implications associated with this philosophical shift in the way school effectiveness was determined that included: society has a right to expect that schools provide students with academic gain regardless of the level in which they arrive at the schoolhouse door and that all students should learn at a level commensurate with their abilities. Amrein-Beardsley & Holloway (2019) contend this shift has resulted in repositioning the role of the teacher into a production function model where teachers produce a product that is consumed by students and is measurable in terms of how much "value" a teacher "adds" to the production. Sanders' mixed-model system, later referred to as the Education Value-Added Assessment System (EVAAS), was developed to minimize/trivialize some of the concerns that were becoming associated with VAMs, such as: missing student data (test scores), various modes of teaching (self-contained, team teaching, departmentalized), teachers changing assignments over the years, transient students, regression to the mean, different variance-covariance structures across school systems, and the need to include concomitant co-variables as needed. The EVAAS model claimed to report measurable differences between schools and teachers with regard to their effect on indicators of student learning (Sanders & Horn, 1998). Long held assumptions were challenged, including a distinction between student achievement and student growth. There was no correlation between school district cumulative growth gains for students receiving free/reduced lunch or for racial composition. Furthermore, teacher effects were not site specific and gain (growth) scores could not be predicted simply by knowing the location of the school (Sanders & Horn, 1994).

Individual states throughout the country soon followed Tennessee's lead. For example, the General Assembly in Indiana passed Public Law 221 (1999), to establish major educational reform and accountability statewide. PL 221 was restructured in 2005, to meet federal

requirements of the reauthorization of the ESEA, known as No Child Left Behind. Specifically, PL 221 outlined criteria that assigned A-F grades to schools (both public and accredited non-public) based on student performance on end of year exams. More recently, the Obama administration implemented *Race to the Top* (U.S. Department of Education, 2009), a competition designed to initiate reform in K-12 public schools. To compete for funding, states had to agree to implement new measures to weight student learning gains as part of teachers' yearly evaluation scores. The U.S. Department of Education's, *A Blueprint for Reform* (2010), called on states and districts to identify effective and highly effective teachers on the basis of student growth. As a result, teacher-level accountability and the "value" teachers "add" to student learning over time was quantified on large-scale standardized tests given at the end of the school year (Barnett & Amrein-Beardsley, 2011). School accountability and individual teacher effectiveness ratings (based on student achievement and growth standards) were measured by calculating a student's actual vs. predicted growth throughout a school year. Layton (2014) reported that 35 states required student achievement to be a significant factor in teachers' evaluations and most were using, or planned to use, VAMs as part of that determination. Collins and Amrein-Beardsley (2014) also reported 88% of states were using, piloting, or developing growth or value-added models to measure teacher effectiveness as a component of teacher evaluations.

Sanders's EVAAS model was the most commonly used VAM in the U.S. (Collins & Amrein-Beardsley, 2014). Studies have demonstrated the model's accuracy at analyzing student academic progress over traditional end-of-year reports focusing on achievement (Amrein-Beardsley, 2008; Collins & Amrein-Beardsley, 2014). The intensity in which student growth was analyzed as a result of VAM use is generally seen as a positive development. As with all

statistical models, there are benefits and limitations that must be considered when evaluating each VAM and its intended use.

Some proposed that The Every Student Succeeds Act (ESSA; 2016) would effectively roll back earlier mandates on teacher evaluation and student achievement and growth (Sawchuk, 2016; Darling-Hammond, Bae, Cook-Harvey, Lam, Mercer, Podolsky & Stosich, 2016). Chenowith (2016) was not so sure as ESSA (2016) accountability standards still include four principles from earlier iterations that suggest change will not happen quickly: (1) the obligation of states to articulate what they expect students to learn; (2) the expectation that schools have an obligation to help all their students meet or exceed standards; (3) the requirement that states assess regularly to measure whether schools are teaching the standards; and (4) the requirement that information about schools, including assessment results, be made available to educators, students, parents, and communities. Furthermore, combining different approaches (classroom observations, student feedback, and value-added student achievement gains) for teacher evaluation, has outperformed traditional measures (Ho & Kane, 2013).

Theoretical Framework

According to Henderikus (2010), an acceptable theoretical framework should provide (a) descriptions of a phenomenon or phenomena, (b) explanations of a phenomenon or phenomena, and (c) predictions about the possible future state of a phenomenon or phenomena in relation to some change. The theoretical framework chosen for this study was Ryan and Deci's (2000) self-determination theory. Ryan and Deci wrote that self-motivation theory was based on "the postulate of three innate psychological needs—competence, autonomy, and relatedness—which when satisfied yield enhanced self-motivation and mental health and when thwarted lead to diminished motivation and well-being (Ryan & Deci, 2000, p. 68). Thus, according to Ryan and

Deci, self-determination is the combined expression of free decision-making (autonomy), the exercise of independent skill (competence), and the feeling of being cared for and valued by others (relatedness).

VAMs can be conceptually explored through each of the components of self-determination theory. Insofar as VAMs might be perceived by teachers as an attempt to control the process and content of teaching, and thereby the freedom of the individual teacher, VAMs might reduce autonomy. As noted in Chapter 1, the drive towards teacher measurement has been closely linked to the need, as expressed in educational policy, to prepare children for higher levels of achievement on various tests, whether the federal tests mandated by NCLB or the state tests in the era of the Every Child Succeeds Act (ESSA) of 2016. To the extent that preparing children for success in high-stakes testing requires a standardization of teaching, and to the extent that the standardization of teaching reduces the freedom of the individual teacher, self-determination theory predicts that the utilization of VAMs would reduce teacher autonomy.

In self-determination theory, autonomy, competence, and relatedness are directly related to the motivation and ability of an individual to engage in a complex task such as teaching (Ryan & Deci, 2000). Therefore, damage to teacher autonomy or to the other bases of self-determination is likely to result in a decrease in both motivation and ability to perform, both of which could, in turn, alter teacher practices. In terms of competence, it was possible that the adoption of a VAM signaled to teachers that their competence will be measured through a single metric, that of student performance, which might not be under the sole, or even the predominant, control of the teacher. To the extent that VAMs reduce the various manifestations of teacher competence into a single, possible unreliable or invalid measurement, self-determination predicts

that the utilization of VAMs would demotivate teachers by signaling that non-VAM-based forms of teacher competence are not important.

Finally, self-determination theory can be applied to VAMs in terms of relatedness. The relatedness component of self-determination reflects the individual need to feel cared for and valued by others (Ryan & Deci, 2000). To the extent that VAMs might make teachers feel that they will be rigidly judged and subjected to close control, VAMs might reduce the feeling of teacher relatedness and thereby do additional damage to teachers' motivation and ability.

As a theoretical framework, self-determination theory can be applied to the relationship between VAMs and teacher practices through a series of possible links between phenomena. The first link in the chain was the utilization of VAMs. The second link in the chain was the likelihood of teachers feeling a loss of autonomy, competence, and relatedness pursuant to the adoption of VAMs. The third link in the chain was the possible decline in teacher motivation and ability as a result of diminished autonomy, competence, and relatedness. The final link in the chain was that decreased teacher motivation and ability could be reflected in other outcomes, including changes in institutional practices. The selection of self-determination theory for the theoretical framework of the study was likely to increase the interpretative and explanatory scope of the findings.

Arguments for VAMs

Sanders (2006) provided a comparison of different classroom-level VAMs and offered pros and cons for selecting a model. Emphasis was placed on VAM development that provided solutions for missing data, poor standard error, teacher comparison against peers (teachers do not have to be winners and losers), and test scaling consistency across grade levels and longitudinal assessments. Sanders recommended a Multivariate Response Model (MRM) due to conservative

results that minimized the likelihood of both false positives and negatives. In this model, a residual or gain score is calculated (student's previous score is subtracted from the current score). By focusing on gain scores, each student serves as their own control for exogenous variables like family characteristics and school influences (Sanders & Horn, 1994).

Sanders (2006) contended the MRM approach provided statistical solutions for scaling, measurement error in predictor variables, missing data, teacher effect scores regressing to the mean, selection bias (all data from each student is used throughout the longitudinal model), individual student scores linked not only to current classroom but all previous classrooms (layering), and accommodations for team teaching (instruction in which multiple teachers are responsible for student learning). In the event, an MRM is not feasible, a Univariate Response Model (URM) was recommended with caution in regards for the model's inability to account for missing data. Using data from tests of the same scale was the primary reason for use of an MRM model, whereas URM models used data from multiple tests with unlike scales (Sanders, 2006).

Other practitioners in the field continued to develop a framework for how to use VAMs. Chudowsky, Koenig, and Braun (2010) concluded that each major class of value-added approaches had shortcomings and there was no consensus on the best approach for VAM use. However, there was enough precision and stability in VAMs to suggest that justification for low-stakes uses would not negatively impact individual teachers, administrators, or students. In addition, VAMs estimates were best used in combination with other indicators, while acknowledging that Sanders' EVAAS model is the 'least bad' of all VAMs. Lipscomb, Chiang, and Gill (2012) also noted that there was a place for VAMs in education, "[VAMs] certainly provide better information for evaluating teacher and school effectiveness when compared against the alternative of maintaining the current system of evaluation in many school districts

and states" (p. 5). Goldhaber and Hansen (2010b) also concurred that VAMs have shown to be a better indicator of teacher effectiveness than teacher graduate degrees, certification, and experience after the initial five years of service. As new models emerged, reliability improved (Goldhaber, 2012; Chetty, Friedman, & Rockoff, 2012) and research indicated that students assigned to higher value-added teachers were more likely to attend college, earn higher salaries, live in better neighborhoods, and save more for retirement (Chetty, Friedman, & Rockoff, 2012). Furthermore, teacher effectiveness (measured by impacts on student test score achievement) was the most important schooling factor when it came to improving student achievement (Rivkin, Hanushek, & Kain, 2005) and a teacher's track record of value-added was the single best predictor of future value-added performance (Kane, 2012).

One point of agreement amongst educational measurement experts was that VAMs should be a part of a multiple set of measures used to identify school or teacher effectiveness (Glazerman et al., 2010; Koedel, Mihaly, & Rockoff, 2015). Unfortunately, the movement to use student achievement and growth metrics for school and teacher accountability was overshadowing the potential use of VAMs as a school improvement tool (Hallinger, Heck, & Murphy, 2014). Furthermore, the level of academic achievement that students attain by eighth grade has a larger impact on their college and career readiness by the time they graduate from high school than anything that happens academically in high school (ACT, 2008). Using the ACT study (2008) as an impetus for change, educational leaders were compelled to ensure that students are maximizing their growth potential and exhausting options to close students' skill gaps. The continued development of value-added models may have also provided additional insight on teacher effects to both achievement and non-achievement outcomes, like motivation, engagement, and grit (Anderman, Anderman, Yough, & Gimbert, 2010; Ruzek, Domina, Conley,

Duncan, & Karabenick, 2014). This approach may have allowed for a more holistic approach to evaluation and by incorporating multiple teachers that students visit throughout the day or year. As a result, due to both the lack of reliability in other models being used for teacher evaluation, and the promise shown in the development of new models in a relatively short period of time, the continued development and improvement of value-added models has been worthwhile (Rivkin et al., 2005; Rockoff, 2012).

Arguments Against VAMs

While VAMs became more and more popular nationwide, there were many educators that suggested continued use was inappropriate due to concerns about oversimplifying the teacher evaluation process on the basis of one-time test scores, the relationship of student background factors on achievement, tracking and estimates for student growth above and below expected ranges, reliability of teacher effect scores, and high stakes testing negatively impacting school morale.

Oversimplification of the teacher evaluation process on the basis of one-time test scores. The oversimplification of the teacher evaluation process, with focus on one-time per year test scores, has been at the forefront of the argument against the use of VAMs (Amrein-Beardsley, 2008; Pianta, 2012). McAnnich (2012) argued that teachers should not be held responsible for how students achieve, but for their intent to teach successfully. Even those that accepted that student achievement should be a part of the evaluation process urged policy makers, state departments, and school districts to not make high-stakes decisions on the basis of a single test score (Amrein-Beardsley & Collins, 2012; Amrein-Beardsley & Holloway, 2019). When making these decisions, multiple indicators, showing similar results, should be used to justify employment decisions (Amrein-Beardsley & Collins, 2012; Braun, 2012). Teachers

preferred a more nuanced evaluation system in which multiple indicators were included. Classroom observations were typically a focal point of these models even though they have demonstrated little explanatory power in predicting student academic outcomes (Weisberg et al., 2009; Glazerman et al., 2010; Kane & Staiger, 2012). The preference for classroom observations is, in part, due to the real-time feedback that is available from administrator to teacher. As a result, teachers and administrators trusted classroom observations more than value-added data (Goldring et al., 2015; Harris & Herrington, 2015; Jiang, Sporte, & Luppescu, 2015). In addition to classroom observations, teachers also contended that credentials, education, and experience should be included in evaluation models (Hanushek, Rivkin, Rothstein, & Podgursky, 2004). In lieu of a shortage of reliable indicators, it was easy to see why policy makers have turned to the use of a statistical method to measure teacher effectiveness.

As recently as 2014, Collins and Amrein-Beardsley conceded that VAMs were likely here to stay. In response to President Obama's *Race to the Top* (2009) initiative, in which states were incentivized to adopt teacher evaluation policies that utilized VAM principles, 88% began implementing VAMs to evaluate teachers. In addition, 30 states were mandated to implement VAM measures by their state legislature (Amrein-Beardsley, 2014). It is yet to be determined what impact ESSA (2016) will have on teacher evaluation and accountability plans. Some states moved forward (Florida, New Mexico, New York, Ohio, Tennessee, Texas), while others (Alabama, Arizona, Georgia, Oklahoma) slowed down and made adjustments that appeared to minimize consequences for teachers that receive a poor VAM score (Amrein-Beardsley & Holloway, 2019).

Part of this was due to states not being far enough along in the development of plans for formative use of growth or value-added data as part of their teacher evaluation or school

improvement processes (Collins & Amrein-Beardsley, 2014). To be effective in a broader context that includes school improvement and professional development for teachers, VAMs must be developed that include a comprehensive set of assessments that all work together with a set of accepted responses to patterns found in assessment data, including short, medium, and long cycle assessments that teachers can use throughout the year to measure improvement (Supovitz, 2009; Pianta, 2012). Until then, basing a teacher's effectiveness on the results from one standardized assessment may change their practice, but the changes will continue to be superficial and focused on content coverage and test preparation (Supovitz, 2009).

Relationship of student background factors on achievement. A point of contention amongst many VAMs critics was the lack of attention given to student background factors (educational attainment of parents, family income levels, individual student ability, race) that are known to impact student achievement (Amrein-Beardsley, 2008; Amrein-Beardsley & Collins, 2012; Diamond & Spillane, 2004; Marchant & Finch, 2016; Konstantopoulos & Borman, 2011; Reardon, 2011). This was supported by landmark studies on the impact of student background factors on student achievement (Anyon, 1981; Coleman, 1966; Lareau, 1987; Rist, 1970). Berliner (2014) effectively summarized this claim by stating the compositional effects, or background factors, that are prevalent in neighborhoods and schools in which they operate are dependent on a series of peer interactions that cannot be measured in a statistical model. The measurement of achievement and growth as two separate entities continued to blur this argument as some were unwilling to accept value-added propositions as worthy endeavors when a number of alternatives might better serve the population at large, including: place public housing in middle class sections of cities, integrate schools by social class, eliminate tracking, establish gender equity in classrooms, balance minorities in classrooms, and establish quality early

childhood programs (Berliner, 2014). Others simply acknowledged that statistical models will never be able to provide fair comparisons of teachers that work in different schools with different students (Haertel, 2013). There were simply too many assumptions that must be accepted to truly measure the isolated impact of teacher effects on student learning (Amrein-Beardsley & Holloway, 2019).

Sanders and Horn (1998) contended that gain scores are not related to racial composition of students, socio-economic level of students, or mean achievement level of the school. This is due to the aforementioned background factors (educational attainment of parents, family income levels, individual student ability, race) known to impact student achievement that are "buried" within the initial test score (Sanders & Horn, 1994). To that end, the best VAMs were based on the notion that analysis of year-to-year growth accounted for influence of student backgrounds that otherwise would impact student test scores. This implied that measuring change or growth, as opposed to achievement, eliminates background factors (racial composition or socio-economic levels of students) that skew achievement data analysis (Sanders, 2006). This challenged long-standing assumptions on student achievement and indirectly forced educators to adopt a growth mindset when evaluating scores in which each student serves as their own control and teachers, schools, and districts that perform the best provide academic growth opportunities for students regardless of prior academic achievement (Sanders & Horn, 1998).

Research began to emerge that supported Sanders and Horn's claim. Wang, Walters, and Thurn (2013) concluded when measuring school growth, the demographic characteristics of students were not relevant. Similarly, Goldhaber and Theobold (2014) and Kane (2012) concluded that as long as baseline achievement is included in the calculation, the inclusion of student demographics and whether or not a student received free or reduced lunch makes little

difference in value-added results. Fagioli (2014) agreed that VAMs do account for student background and cited several low-SES schools in a California study that showed remarkable performance, scoring well above expected levels. In the same study, high-SES schools and advanced students were not negatively impacted. This was important as critics of VAMs contended schools with large percentages of high achieving students did not show positive growth, relative to schools with lower achieving students, due to restrictions at the top end of the scale (ceiling effect) used to measure results (Koedel & Betts, 2010).

Fagioli's study (2014) demonstrated the advantages of using a value-added approach versus the state accountability metric that was used in California. The state model exhibited a high correlation between SES and race with school accountability ratings. Sacks (1999) referred to this phenomena as the "Volvo Effect:" one can guess a school's performance by looking at the make and model of the cars in the school parking lot. When using a VAMs approach, growth was measured after establishing year-to-year targets based on how a student arrives at the schoolhouse door. This was consistent with what Sanders and Horn (1994) claimed all along, that school gains are not correlated with racial composition or the percent of students receiving free or reduced lunch. Today, some of the more robust models now control for student-level factors (student SES, school SES, ethnicity, parent educational attainment, tracking), with error rates at acceptable levels (Schochet & Chiang, 2010; Parsons, Koedel, & Tan, 2019).

Tracking and estimates of student growth. The U.S. education system has long been one of the most segregated and unequal in the industrialized world (Darling-Hammond, 2015). One of the contributing factors has been the practice of tracking, or within-school ability grouping (Ladd, 2012). This refers to any school practice that increases the homogeneity of instructional groups by stratifying students by curriculum standards, educational career goals, or

ability (Akos, Lambie, Milsom, & Gilbert, 2007). Even though tracking has been widespread, most VAMs have not taken into account this important school-level factor (Dieterle, Guarino, Reckase, & Woolridge, 2012; Hallinger et al., 2014; Morganstein & Wasserstein, 2014). Some argued that it would seem logical for teachers to be compared with other teachers that are teaching similar students and the failure to do so would create bias (Dieterle et al., 2012; Morganstein & Wasserstein, 2014; Rothstein, 2009). Questions about bias were also raised by Pathak et al. (2017) and Angrist et al. (2016) while studying conventional VAM estimates from admissions lotteries in Boston and Charlotte. When assignment mechanisms were controlled for, teacher rankings were shown to have a positive correlation and a low rate of misclassifying teachers as high- or low-performing (Guarino, Reckase, & Woolridge, 2012). As a result, multi-layered models that accounted for variance in student achievement and non-random assignment of students were likely to become more common (Hallinger et al., 2014).

In addition to uncertainty over value-added models designed to account for nonrandom distribution of students, teachers also felt as if they were punished for teaching students at both the low (special education and English-language learners) and high end (honors and gifted and talented) (Amrein-Beardsley, 2012; Amrein-Beardsley & Collins, 2012; Collins, 2014; Darling-Hammond, 2015; Darling-Hammond, Amrein-Beardsley, Haertel & Rothstein, 2012; Pianta, 2012; Misco, 2008). Furthermore, some teachers felt that it was difficult to show growth in back to back years when looping with students (Collins, 2014). To make matters worse, many value-added models did not have solutions for the 'ceiling effect.' This phenomenon occurs when grade level exams do not allow students to show gains that exceed expectations due to only measuring grade-level standards without including items from earlier or later grade levels

(Amrein-Beardsley, 2012; Darling-Hammond, 2012; Darling-Hammond, 2015; Koedel & Betts, 2010).

An additional concern about the nonrandom placement of students was that this was most likely to occur during a student's high school years. Most value-added studies have been conducted at elementary and middle schools (Rockoff, 2012; Johnson, 2015). Making teacher effect determinations for high school teachers, in which students may spend the day with as many as seven different teachers, has been next to impossible (Hallinger et al., 2014; Jackson & Bruegman, 2009). It was not surprising that there were reports of teachers with the opportunity to change teaching assignments becoming savvy about moving out of grades and subject areas where it was tough to show growth (Holloway-Libell, Amrein-Beardsley, & Collins, 2012).

Reliability and instability. There have been a number of VAMs that have produced misleading results (Glazerman et al., 2010; Sanders, 2006). The arguments ranged from instability due to exogenous variables (Berliner, 2014) to imprecision based on estimation errors (Ballou & Springer, 2015). Researchers were in agreement that there was potential for misclassifying high- and low- performing teachers due to the high stakes context of public education (Guarino et al., 2012). The Houston Independent School District inappropriately used inconsistent data to make high stakes decisions about teachers that resulted in four teachers being terminated for poor value-added scores (Amrein-Beardsley & Collins, 2012). Additional concerns about use of value-added data were related to teachers not understanding value-added metrics well enough to improve instruction and mixing norm- and criterion- referenced tests together to produce results. In other parts of the U.S., researchers reported that error rates as high as 25% were still prevalent even with three years of longitudinal data being used in the value-added calculation (McCaffrey, Sass, Lockwood, & Mihaly, 2009; Schochet & Chiang,

2010). It was no surprise that Ballou and Springer (2015) came to the conclusion that if teachers really knew how much error is associated with most value-added models, their level of confidence in the measures would decrease even more. Amrein-Beardsley & Holloway (2019) contended that teachers work in highly complex institutions that are subject to uncontrollable factors and to try and reduce teaching to a single numerical outcome is indefensible when taking into account all the literature on the topic that suggests otherwise.

Sample size and testing conditions. There were also concerns about the reliability of teacher estimates when only dealing with a handful of students. As a result, value-added scores for teacher-level analyses were subject to a considerable degree of random error when based on the amount of data that are typically used in practice for estimation (Schochet & Chiang, 2010). Even as estimates became more reliable when combining multiple years of data, teachers or schools with fewer students in the data set regressed to the mean over time (McCaffrey et al., 2009). Additional school level effects that impacted the reliability of value-added scores included teachers verifying their own rosters and being less likely to claim students with lower scores (Amrein-Beardsley, 2008; Ballou & Springer, 2015), students more likely to score higher when their own teacher administers the test (Ballou & Springer, 2015), and teachers reporting thoughts of cheating or teaching to the test (Collins, 2014; Haertel, 2013).

Error rates. The American Statistical Association (2014) reported that teachers accounted for 1-14% of the total variation of student growth and Darling-Hammond (2012) indicated that the teacher effect increased to 4-25% of the variation with more than two years of data. Using either measure, teacher effects accounted for less than a quarter of the total variation in student growth. McCaffrey et al., (2009) concluded this is problematic especially when correlations did not have an acceptable level of reliability. Most estimates placed teacher

effectiveness error rates between .30 and .50 (Corcoran, 2012; Goldhaber & Hansen, 2010a; Rockoff, 2012). Educators should consider examples from other sectors of the economy in which imprecise measures are used to make high stakes decisions (McCaffrey et al., 2009; Glazerman et al., 2010). Examples could include, use of the SAT in college admissions decisions, year-to-year correlations in patient mortality rates within the medical field, home sales for realtors, returns on investment funds, and baseball batting averages as ballplayers negotiate new contracts (Glazerman et al., 2010). As with the aforementioned examples, value-added model error rates became much more stable when including multiple years of data (Koedel & Betts, 2010; McCaffrey et al., 2009).

Validity studies and policy implications. A universally accepted model has not yet been carefully tested (Amrein-Beardsley, 2008; Morganstein & Wasserstein, 2014). This is due in part to many different models in use that control for different factors. As a result, data from the same school can produce different value-added results depending on the variations within the model (Morganstein & Wasserstein, 2014). This not only increases the pressure for a solution on a national level, but intensifies the claim that VAMs should not be used for important decisions on the classification of teachers (Collins, 2014; Darling-Hammond, 2012; Guarino et al., 2012; Ladd, 2012; Pianta, 2012; Schochet & Chiang, 2010). As if that were not enough, additional concerns have surfaced about the instability of VAM scores over time due to regression to the mean (Darling-Hammond, 2015; Rothstein, 2012). The call to discontinue the use of VAMs has been getting louder. Hallinger et al. (2014) urged policy makers to quit wasting time and money on the development of value-added strategies "due to the efficacy of instructional leadership and school improvement strategies not meeting the dual criteria of empirical evidence and feasibility" (p. 19). Yet, as of January 2019, 48 states (California and Kansas being exceptions)

are still rating schools in part on student growth based on test score results to comply with ESSA requirements (Data Quality Campaign, 2019).

Morale and unintended consequences. Use of value-added measures could have the potential to do even more harm than actively criticizing teachers (Ewing, 2011) due to negatively impacting collaboration and trust amongst teachers and administrators (Haertel, 2013; Johnson, 2015). In the most successful schools, teachers worked together effectively (Little, 1982; Fullan, 2000). Haertel (2013) proposed that teachers may be less likely to help one another for fear of reducing their own value-added measure. While that remains to be seen, teachers have reported a sense of heightened pressure and competition that results in poor morale and puts a strain on collaboration (Amrein-Beardsley, 2008; Collins, 2014; Jiang et al., 2015). The potential impact of driving teachers away from schools that need them most should not be overlooked. Since 2000, turnover rates in high poverty schools have been much higher when compared with schools in higher income communities (Ingersol, 2001). This turnover can devastate learning communities within schools and take years to rebuild. Due to these unintended consequences, teachers should not be ranked by VAM scores (ASA, 2014).

Lessons from the past. While it is not necessary to focus on educator perspectives, perceptions do shape behavior (Harris & Herrington, 2015) and there is evidence that policy implementation rarely impacts practice (McLaughlin, 1987; Wilson, 2003). In addition, the most motivated teachers make up a minority of the workforce (Elmore, 1996). This suggests that only one quarter of the profession is intrinsically motivated to question their practice on a fundamental level and look to outside models to improve teaching and learning. This sobering thought should not make anyone think that accountability demands will lead to lasting change within the profession.

Statistical intimidation. Data analysis on large-scale assessments is useful to measure school or system progress, but has been limited for instructional guidance at the classroom level (Supovitz, 2009). Teachers have wanted to know what they are doing well and what they need to improve upon. This kind of feedback is typically done through classroom observations and in professional learning communities. Value-added measures typically occur only once per year with little time to adjust teaching strategies (Harris & Herrington, 2015). Because of the emphasis on these one-time per year tests, teachers have reduced time spent on important content and skills (Amrein-Beardsley, 2008; Darling-Hammond & Adamson, 2014; Jiang et al., 2015). In addition to a lack of feedback and a reduction in time spent on task, many value-added models were very technical and difficult to understand for those not in the field (Ewing, 2011). To complicate matters even further, teachers and administrators are likely to need a background in statistics to understand most VAMs (Amrein-Beardsley, 2008), while in the process turning education into a numbers game (Weingarten, 2014).

Review of Empirical Evidence

The purpose of this sub-section of the literature review is to examine evidence related to changes in instructional practice as a function of the introduction of VAMs. This sub-section of the literature review is directly related to the research questions of the current study, which focus on the measurement as well as the exploration of possible changes to teacher practice that might be attributable to the adoption of VAMs. Key studies are described and discussed individually, then related to each other in the next sub-section of the literature review, which contains an identification of points of consensus related to the current and possible future state of VAMs and their impact on teacher practices.

A study carried out by Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012) reached the conclusion that many teachers in Houston had not changed their teaching practices despite becoming subject to a VAM, the Education Value-Added Assessment System (EVAAS). Darling-Hammond et al. quoted a Houston teacher as stating the following:

> I do what I do every year. I teach the way I teach every year. [My] first year got me pats on the back; [my] second year got me kicked in the backside. And for year three, my scores were off the charts. I got a huge bonus, and now I am in the top quartile of all the English teachers. What did I do differently? I have no clue (Darling-Hammond et al., 2012, p. 11).

On the basis of evidence of this kind, Darling-Hammond et al. concluded that the implementation of EVAAS and other VAMs was not necessarily associated with changes in teaching practice. Darling-Hammond et al. suggested that teachers were more or less stable in their instructional practices, regardless of whether or not they were subject to VAMs. Darling-Hammond et al. did not uncover evidence of teachers altering their teaching practices in response to a VAM.

However, Darling-Hammond et al.'s (2012) study was flawed in two respects. First, Darling-Hammond et al. did not attempt to systematically define or measure teaching practices; teachers were at liberty to describe, or not describe, their own instructional practices. Second, Darling-Hammond et al. did not apply a pre-test / post-test approach to formally measure changes in teaching practice as responsive to VAM onset. Darling-Hammond et al. asked teachers about their teaching practices without strictly sorting these responses into before and after categories that could have been subsequently explored through either quantitative or qualitative means.

The connection between VAMs and changes in teacher practices was made more explicit in the work of Boyland, Harvey, Quick, and Choi (2014) in their discussion of VAMs in Indiana. In their review of literature, Boyland et al. noted that VAMs created both a judgment and accountability structure according to which there would be a basis to expose teachers to certain forms of professional development if their performance on VAMs was low. However, Boyland et al. did not address the questions of (a) how frequently teachers might be subject to professional development or other attempts to change their instructional practices, and (b) to what extent teachers might actually change their practices in response to such measures undertaken by principals.

The conclusion reached by Darling-Hammond et al. (2012) was twofold, namely that (a) teachers are fairly fixed in their instructional habits and (b) teachers are skeptical that chasing VAMs by altering their own instructional practices will be effective. To these possibilities, Boyland et al. (2014) added the possibility that teachers who do not perform well in an era of VAMs will have limited voluntary scope to change their own practices and might instead be compelled to change. However, as is the case in Darling-Hammond et al.'s work, Boyland et al. did not attempt to systematically define teaching practices, strictly time before- and after-VAM effects, or, in the context of their systematic review, present findings related to actual instructional practice changes related to principals' VAM-driven decisions about the necessity for teacher development.

Darling-Hammond and Youngs (2002) reviewed the programmatic logic behind the relationship between VAMs and changes in teacher practices by means of a review of the 2002 Secretary's Annual Report on Teacher Quality published by the United States Secretary of Education. Darling-Hammond and Youngs began by acknowledging the figure, given in the

Secretary's Annual Report on Teacher Quality, that 7% of variation in student test scores can be ascribed to teachers, but also noted that, according to the evidence reviewed by the United States Secretary of Education, the potential value added by a teacher might not be the result of instructional practices themselves. Bondy, Ross, Hambacher, and Acosta (2013, p. 27) identified 'warm demanding' as a teacher orientation more likely to result in student improvement than teaching practices, noting that 'warm demanding' required the teacher to:

- Convey personal interest in students
- Understand how students perceive care (for example, in varying cultural contexts) and ensure that demonstrated care is communicable and meaningful to the student
- Work hard to remain committed to the welfare of students
- Ensure that ordinary actions (such as pedagogy, gestures, classroom management, etc.) are utilized at least as frequently as verbal affirmation in order to convey care
- Be insistent in bringing students back from digressions to their assigned work
- Put academic work before 'fun' activities and breaks
- Remind students that they are capable of high achievement
- Hold students accountable for high achievement

Regardless of whether teacher practices or teacher orientations (such as 'warm demanding') are more likely to correlate with student achievement, the logic of VAM-based professional development is that (a) VAM can identify teachers in need of professional development, and (b) professional development can lead to improvements in practice or orientation that then lead to improvements in student performance. This logic was implicitly acknowledged, but not empirically tested, by Darling-Hammond and Youngs (2002), Boyland et al. (2014), and Darling-Hammond et al. (2012). However, research carried out by Springer et al.

(2010) on the basis of an experiment in Tennessee suggested that, even if VAMs can sort teachers into different achievement categories, instructional practices are not likely to change in response to measurement.

Specifically, Springer et al. (2010) reported on the results of an experiment carried out in the state of Tennessee in which teachers were promised an annual bonus of $15,000, a considerable premium to their actual salaries, if they were able to raise student performance. As Springer et al. argued, the strong monetary incentive, combined with the voluntary nature of this experiment, meant that the teachers who decided to participate were strongly motivated to change various aspects of their teaching practices, personal orientations, and other variables in order to obtain the reward. However, the result of the multi-year experiment in Tennessee was a failure. Teachers who entered the bonus-eligible group were not only unable to increase the performance of their students to significantly higher percentiles; their students were statistically indistinguishable from the students of teachers who did not elect to participate in the monetary experiment.

One of the conclusions drawn by Springer et al. (2010) was that teachers are resistant to changing their practice, especially in the short-term and on the basis of an external motivation. In this context, one point of interest in the Tennessee experiment was that the monetary bonus, voluntarily sought by the teachers, was a positive motivator, whereas, as noted in Chapters 1 and 2, VAMs have been perceived as a form of punishment or at least unwanted scrutiny or pressure. In theory, then, the availability of a reward should have created an incentive for teachers to alter their teaching practices, but, as Springer et al. found, such practices remained remarkably stable over the experimental period. Springer et al.'s qualitative interviews generated some evidence that teachers were in fact trying to alter their practices so as to increase their students' test

performance; however, despite these efforts, instructional practices remained remarkably stable among the sample of teachers.

On the basis of self-determination theory (Ryan & Deci, 2000), motivation and energy result from an encouragement of autonomy, competence, and relatedness, whereas demotivation and burnout result from a curtailment of these basic psychological needs. It is possible that instructional practices in Springer et al.'s (2010) study of the Tennessee experiment did not change because motivation and energy are not necessarily reflective of changes in instructional practices. Instructional practices might be based not on changes in motivation and energy, which are transient, but on deeply embedded knowledge, complex behaviors, habits, and other variables that are difficult to alter, even if a VAM furnishes a punitive mechanism for teachers who fail to alter their teaching habits.

One plausible means of analyzing the relationship between VAM implementation and teaching practices is through a framework in which VAM environments are compared to pre-VAM environments. However, as noted by Weisberg, Sexton, Mulhern, and Keeling (2014), another possible research approach to this topic is through an analysis of differing types of VAM environments. As Weisberg et al. noted, some VAM environments are solely punitive (that is, focused on utilizing the results of VAMs for either teacher remediation or teacher dismissal) whereas, in other VAM environments, VAM results might also be utilized for positive as well as punitive purposes. A comparison of instructional practices in positive and punitive VAM environments would also be a means of measuring and exploring the kind of influence VAM might have on teaching practices. However, as Weisberg et al. noted, no such analyses appear to have been carried out by empirically oriented researchers.

Another possible reason that VAMs might not prompt changes in teaching practices was noted in Shugart's (2017) dissertation. In several cases, Shugart noted that teachers who were subject to VAM scrutiny believed both that (a) they had to put on a show for observing principals and (b) there was no guidance about what kind of instructional practice changes were actually necessary in order to improve their assessment. In terms of point (a), Shugart quoted three different teachers:

> You put on a show while they're in here, they walk out, you go back to what you normally do….I am a good teacher, but the moment they say it is time for us to come around, I feel it's almost a dog and pony show every day….There was a time that I would plan my lessons when I knew they were coming in and I had the opportunity to plan…(Shugart, 2017, pp. 87-89).

Shugart interpreted the quotes to mean that teaching practices remained relatively stable even when teachers understood that they were scrutinized as part of a VAM. The teachers in Shugart's study were all serving at a school that had just implemented a VAM, but, despite their knowledge of the stakes involved, the teachers were far more concerned with demonstrating that they were active in the classroom rather than changing their own teaching practices. The teachers did not indicate any beliefs about the necessity of changing their teaching practices.

Point (a) is of interest as a further explanation of the durability of teaching practice; the teachers in Shugart's study appeared to believe deeply in their own practices but lacked confidence that these practices would be noted and appreciated by observers in the context of a VAM. Point (b) is of interest as a novel explanation for why teaching practices might not change in response to VAM; in the absence of appropriate guidance, teachers might simply not know which instructional practices to change or how to go about changing them.

VAMs: The Current and Future State

The literature reviewed in this chapter indicates that VAMs are widely implemented and used by policy makers, but not trusted by teachers and educational experts. One of the reasons for the status quo of VAMs, insofar as educational leaders and policy-makers are concerned, is that the public sector is expected to rely on measurements and data-driven approaches to accountability (Weisberg et al., 2014). Although measurements such as those produced by VAMs might be flawed, they have the possible advantage of giving decision-makers easily understood numbers that can be utilized for decision-making. However, even among VAM supporters, VAMs have been critiqued for their support of remediation and firing decisions more than of other kinds of decisions (Weisberg et al., 2014). Among teachers, VAMs appear to be deeply resented, mainly because of the belief that VAMs are an unreliable and invalid mechanism for measuring teacher performance; some of the available statistical support for this claim was noted earlier in this chapter.

The future state of VAMs is likely to be driven by the performance trends in American public education. As noted in both Chapters 1 and 2, VAMs were developed because of the perceived need to improve the academic performance of American students. On the assumption that teachers are extremely important contributors to such improvement, and that teacher performance can be positively influenced, VAMs provide a means of determining whether teachers are in fact improving. The programmatic logic of VAMs is that (a) teachers with performance problems can be identified, (b) teacher performance problems can be remediated, and (c) increased teacher performance will lead to improved student performance. However, the future viability of VAMs is somewhat threatened by the fact that American students' performance in mathematics and reading has been flat. If these trends continue, one possible

conclusion is that possible teacher contributions to student improvement might be limited by structural factors, that is, factors not related to teachers and thereby out of the control of teachers. Figure 1 below, based on data from the National Assessment of Educational Progress (NAEP, 2017), indicates that performance in grade 4 mathematics has been statistically indistinguishable from 2017 levels since 2007; in other words, grade 4 students have not obtained a significant improvement in mathematics performance over the past decade.

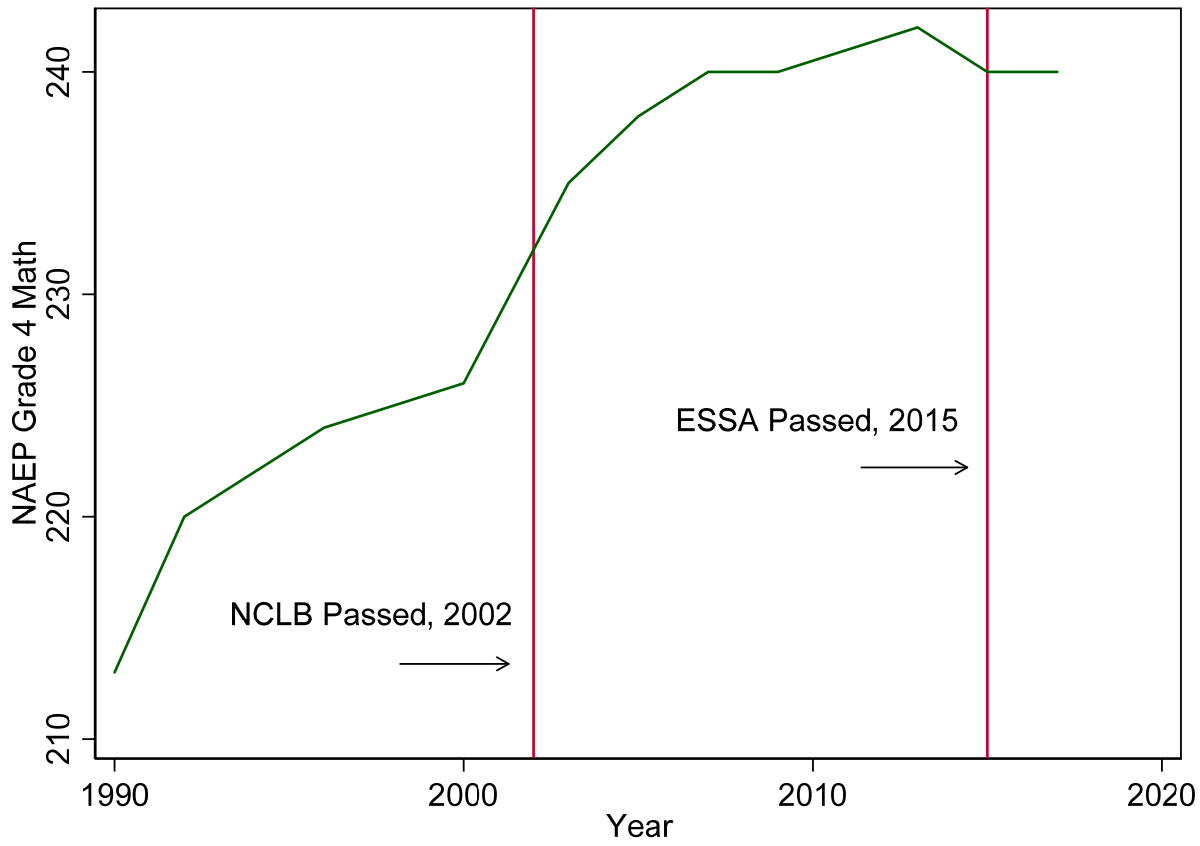

Figure 1. Annual change in grade 4 mathematics performance, NAEP. *Note*: Original figure based on tabular data from NAEP (2017).

Stagnation in performance might intensify demands for the use of VAMs and similar measures to accentuate accountability or, conversely, lead researchers and policy-makers to conclude that VAMs cannot generate the future improvements that are necessary for American students' performance.

The trend observed in grade 4 mathematics also exists in grade 8 mathematics. In terms of NAEP's (2017) measurements, grade 8 mathematics performance stagnated in 2009; in other words, performance from 2009 onwards has been statistically indistinguishable from the 2017 figure, indicating that American eighth graders have not significantly improved their mathematics performance in nearly a decade.

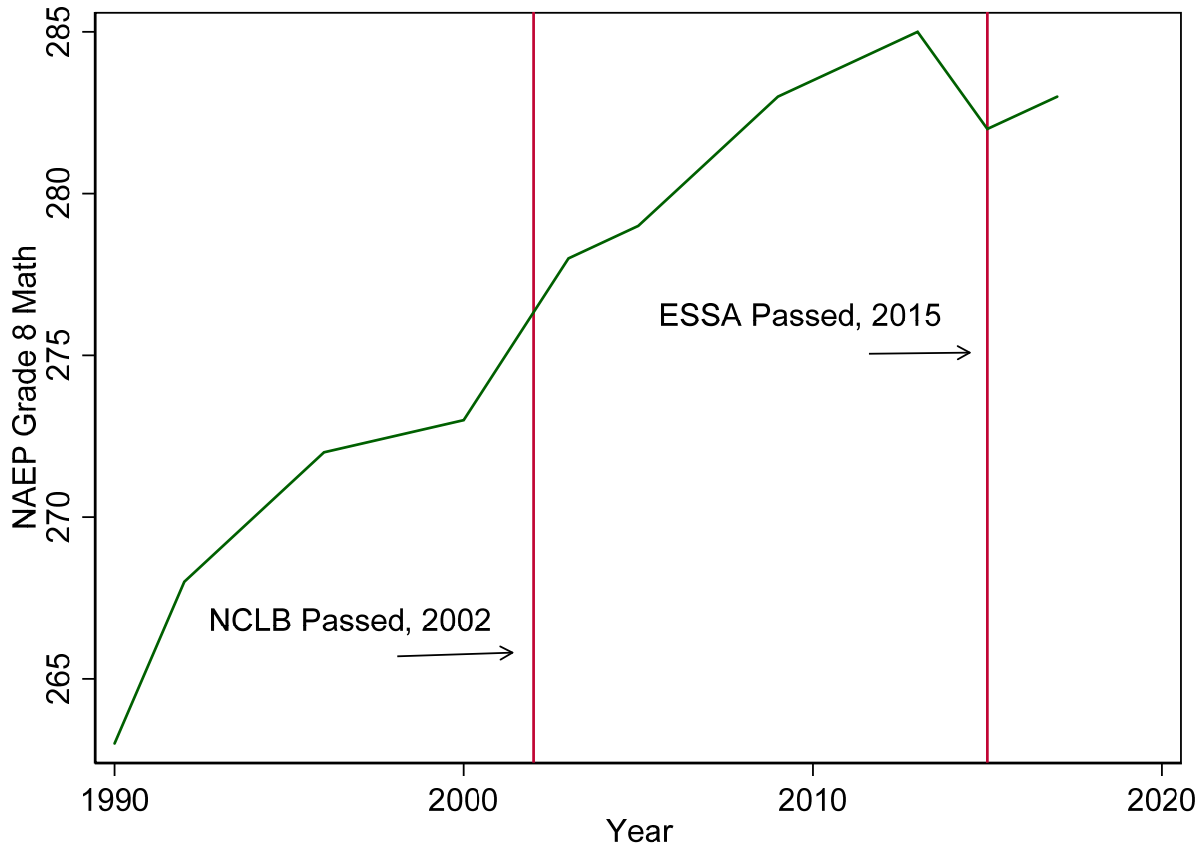

Figure 2. Annual change in grade 8 mathematics performance, NAEP. *Note*: Original figure based on tabular data from NAEP (2017).

The patterns charted for grade 4 and grade 8 mathematics also appear in grade 4 and grade 8 reading, and in grade 12 reading and mathematics. In each case, NAEP data indicate that student performance has been stagnant for about a decade. An evaluation of these longitudinal trends suggests the possibility of either intensification of reliance on VAMs (as a form of

bringing more accountability to bear on teachers) or the formation of a new consensus that VAMs might not be appropriate to positively influence teacher performance.

Gaps In The Literature

One of the most important gaps in the literature on VAMs is the absence of empirical studies that have attempted to measure the relationship, if one exists, between the use of VAMs and the alteration of teacher practices. Outside of the work done by Hamilton et al. (2013), there has been minimal research done on the actual content of teaching practices and impact on student performance as measured by accountability metrics. Items on ATI-R (Trigwell et al., 2005; Appendix A) and scoring devise (Appendix B) represent one plausible approach to systematically identifying, and analyzing changes to, specific elements of teacher practice. Without the use of a systematic scale of this kind, it is not clear that researchers have the ability to track changes in teaching practices.

A second gap in the literature is the absence of pre-test / post-test approaches. On the assumption that VAM implementation presents a possible impetus for teachers to alter their teaching practices, only a pre-test / post-test approach is capable of detecting whether teaching practices (of the kind categorized in Appendix A) might have changed significantly in the wake of VAMs. For this reason, additional research is needed to quantify as well as explore the phenomenon of instructional practice change in response to VAM implementation in American schools.

Conclusion

As state departments of education and local school districts have adopted statistical models to evaluate student performance, VAMs have been used to measure student achievement and growth. Furthermore, by focusing on student outcomes, teachers and schools have been free

to use whatever methods prove practical in achieving student academic progress. This has resulted in the development of programs that appeal to interest groups on both sides of the political aisle. Those committed to equality of education opportunity have been given a centralized plan. Those believing that competition is needed to improve our schools have been given national testing and high-stakes accountability. The development of a mathematical tool to measure student and school performance was a logical next step for those needing to quantify the successes and failures of the reform movement. To that end, the United States educational system has evolved into one of the most decentralized of all industrialized nations (Kober & Usher, 2012). Unlike most G-8 nations, the U.S. relies on state mandated standards of proficiency for school and district accountability. As students compete for jobs in a global society, educational leaders must continue to search for tools that promote student achievement and growth. While questions remain about the accuracy and stability of VAMs estimates of schools, teacher, and program effects for school accountability purposes, the development and use of VAMs estimates for school improvement purposes is appropriate.

This chapter of the study contained an overview and discussion of some of the major themes in the current scholarly literature on VAMs, accountability, and high-stakes testing in American education, and established a context through which the impact of VAMs can be understood. The research, methods, and design described and defended in the third chapter of the study address the gaps in the literature by proposing a mixed-methods, case study to explore the relationship between teacher practice before, and after, VAM implementation in a particular school.

CHAPTER 3

METHODS

In this chapter, the research methods and design of the study are described and defended. In order to achieve this purpose, the chapter is structured as follows. First, the research methods of the study, that of mixed methods, is identified, described, and justified with particular reference to a case study format. Second, the research design of the study, the case study, is identified, described, and justified. Third, the research questions and hypotheses of the study are presented. Fourth, the setting, sample, and population of the study are described. Fifth, data collection is described. Sixth, data analysis is described and justified for each of the research questions of the study. Seventh, the reliability, validity, trustworthiness, and generalizability of the study are discussed. Eighth, the limitations, delimitations, and assumptions of the study are described. Ninth, the ethical issues pertaining to the study are identified. Tenth, positionality is discussed. Finally, all components of the third chapter are briefly summarized as part of a transition to the fourth chapter of the study.

Purpose Statement

The purpose of this mixed-methods, case study was to examine the impact of the adoption of a single school's VAM on the practices of teachers. The presumption of the study was that the imposition of the VAM was not likely to have created changes in the practices of teachers. The purpose of the study will be achieved by answering two research questions: (1) What changes in teacher practices have occurred since the administration of a VAM in a particular school? (2) Did VAM alter the practices of teachers? Why or why not?

Research Questions And Hypotheses

The quantitative research question and accompanying hypotheses of the study were as follows:

RQ1: What changes in teacher practices have occurred since the administration of a VAM in a particular school?

H1_0: Teacher reports of their practices before the administration of a VAM in a particular school were equivalent to teacher practices after the administration of a VAM in a particular school.

H1_A: Teacher reports of their practices before the administration of a VAM in a particular school were not equivalent to teacher practices after the administration of a VAM in a particular school.

The qualitative research questions of the study were as follows:

RQ2: Did VAM alter the practices of teachers? Why or why not?

Research Methods

The research methods utilized in the study were explanatory mixed methods (Creswell & Plano Clark (2011), and a case study format was used. Quantitative research methods have been described (Balnaves & Caputi, 2001; Berger, 2013; Bernard & Bernard, 2012; Davies & Hughes, 2014; Duffy, 1987; Venkatesh, Brown, & Bala, 2013) as particularly appropriate when the underlying research problem involves the measurement of the impact of one or more variables (predictors or independent variables) on one or more other variables (outcome or dependent variables); whereas qualitative methods are more appropriate when exploring contexts, reasons, and underlying explanations based in subjective interpretation and experience (McNabb, 2010). The research problem discussed in the first chapter of the study was based on an absence of knowledge about the impact of a variable (the adoption of VAM in a school) on the practices of

teachers in that school. Because the research problem of the study was intrinsically connected to measuring the impact of a variable on other variables, quantitative research methods were appropriate for the study. However, quantitative methods cannot address issues of how and why teachers might have changed their practices in response to the VAM; such an examination required a qualitative approach.

Teachers' practices have been understood from a variety of perspectives (Bonner, 2014; Collinson & Tourish, 2015; Crawford, 2017; Griner & Stewart, 2013; Stevenson, 2013), with researchers suggesting that such practices constitute what, in another context, McNabb described as "multiple, subjectively described realities" (McNabb, 2010, p. 225). Teachers' practices are both personal and unique to themselves. However, in post-positivistic approaches (Bernard & Bernard, 2012; Creswell, 2015; Jackson, 2015; Trochim, Donnelly, & Arora, 2015) to quantitative methods, researchers have made the assumption that irreducibly complex, subjective, and unique attitudes can be measured and incorporated into quantitative models. In post-positivism, the use of mathematical measurement is only a tool of convenience and a methodological assumption, not a statement about the subjective or objective reality of phenomena such as, in the case of this study, teacher practices. The post-positivistic approach applies to the quantitative rather than the qualitative portion of the study. In keeping with the approach known as descriptive phenomenology (McNabb, 2010) and employing techniques used in explanatory mixed methods (Creswell & Plano Clark, 2011) research during the qualitative part of the study, participants were asked how and why they altered their practices in response to the VAM.

The scope of the methods can also be understood in terms of a case study format. According to Yin (2009), the case study format is particularly suitable to the study of a

phenomenon or research problem "in depth and within its real-life context" (Yin, 2009, p. 18), and the case study format is particularly justified when "the boundaries between phenomenon and context are not clearly evident" (Yin, 2009, p. 18). The VAM utilized in this study was developed by the school. The effect, or lack of effect, of this VAM on teachers within the school is not necessarily likely to reflect national trends. In addition, the study is delimited to the context of a single school, lowering the likelihood that the results will generalize to other settings. For these reasons, the quantitative methods to be applied in this study can be considered as belonging to a case study format.

Research Design

In the tradition of quantitative methods, there are four major approaches to research design. One of these approaches, the correlational approach, has been described in the following manner:

> The variables included in correlational research are isolated and measured by the investigator, but they are characteristics that occur naturally in the subjects…a correlation study consists of establishing a relationship between variations in the *X* variable to variations in the *Y* variable (Keppel, Saufley, & Tokunaga, 1992, p. 460).

In this study, there was an *X* variable (VAM) and a *Y* variable (teachers' and their practice orientation), but the existence of VAM was not naturally occurring. The VAM that is the focus of this study was developed and implemented by the school; in this sense, it represents what methodologists refer to as an intervention (Cassell & Symon, 2004; Creswell, 2015; Creswell & Plano Clark, 2011; Hesse-Biber, 2012; Jackson, 2015; Leary, 2011; McBurney & White, 2011; Trochim et al., 2015; Zikmund, 2003).

All of the teachers at the school that were teaching courses in subjects measured by a VAM when implemented during the 2014-15 school year, and still teaching the same courses during the 2017-18 school year, were invited to participate in the study. There were 55 teachers that met the criteria. As a result, the study design was a case study (Yin, 2009).

The qualitative design was based on descriptive phenomenology (McNabb, 2010). In descriptive phenomenology, participants are asked to discuss their own experiences of a phenomenon (McNabb, 2010). The phenomenon of interest in this study is a VAM, and the descriptive phenomenological design was based on analyzing both how and why teachers at the school changed practices and attitudes as a result of the VAM's imposition.

Setting, Sample, and Population

The population of the study consisted of teachers in the United States who have been impacted by VAMS and work in high achieving, suburban high schools. However, in keeping with the case study format described earlier in the chapter, the study was delimited to a single setting. The setting for the study was a suburban high school in Northwest Indiana with approximately 2,750 students. Prior to the study, the school had received the highest distinction from the Indiana Department of Education, Four Star School, for the past seven years. The school had a history of meeting high quality standards and implementing a continuous process of improvement. To that end, the North Central Association Commission on Accreditation and School Improvement (NCA CASI), an accreditation division of AdvancED, awarded the school their highest distinction during accreditation visits in 2014 and 2019. The sample consisted of 55 teachers. These teachers were teaching courses in subjects measured by a VAM when implemented during the 2014-15 school year and still teaching the same courses during the 2017-18 school year. Out of the 55 teachers eligible, there were 26 males and 29 females. The mean

age of all eligible teachers was 40.74 with a range between 27-62 years of age (beginning of the 2017-18 school year). The average number of years of experience for teachers was 15.36 with a range between 4-39. There were 16 English teachers, 16 mathematics teachers, 12 science teachers and 11 social studies teachers eligible for the study. Surveys were distributed to all teachers in the sample. Age, gender and number of years' experience were not documented on an individual basis to protect anonymity. Teachers were asked to identify their subject area as this information did not reveal the teacher's identity, but assisted in development of interview questions used in the qualitative part of the study.

Surveys were delivered to study participants by email. All participants were asked to complete their instrument on the basis of both their current (post-VAM) practices but also on the basis of their practices, as best as they could be remembered, for the pre-VAM era. Each instrument had questions organized to capture pre-VAM and current (post-VAM) practices. An *a priori* sample size calculation was carried out on the basis of the research questions of the study. Each research question involved a repeated-measures comparison, that is, a comparison of teacher scores on the survey instrument before and after the school's implementation of a VAM. Because participants were compared to themselves, before and after implementation of a VAM, the statistical format of the study was that of a paired-samples *t*-test, which was used as the basis for the *a priori* sample size calculation of the study.

Cohen suggested (Cohen, 2013) that, for an *a priori* sample size calculation in a paired-sample *t*-test approach, the Alpha should be .05, and the statistical power should be .80. In addition, Cohen recommended the adoption of a medium effect size of 0.50. Finally, the hypotheses of the study were expressed in a two-tailed format, a decision that impacts the *a priori* sample size calculation. These inputs were entered into G*Power statistical software

(Faul, Erdfelder, Buchner, & Lang, 2009) in order to generate an *a priori* sample size for the study. The recommended sample size was 34 (see Figure 1 below for the G*Power calculation). Therefore, given the distinct sub-samples of the study, it was appropriate to sample, at a minimum, 34 teachers. A *post hoc* sample size calculation is provided in chapter four based on the final number of personnel for whom data were available.

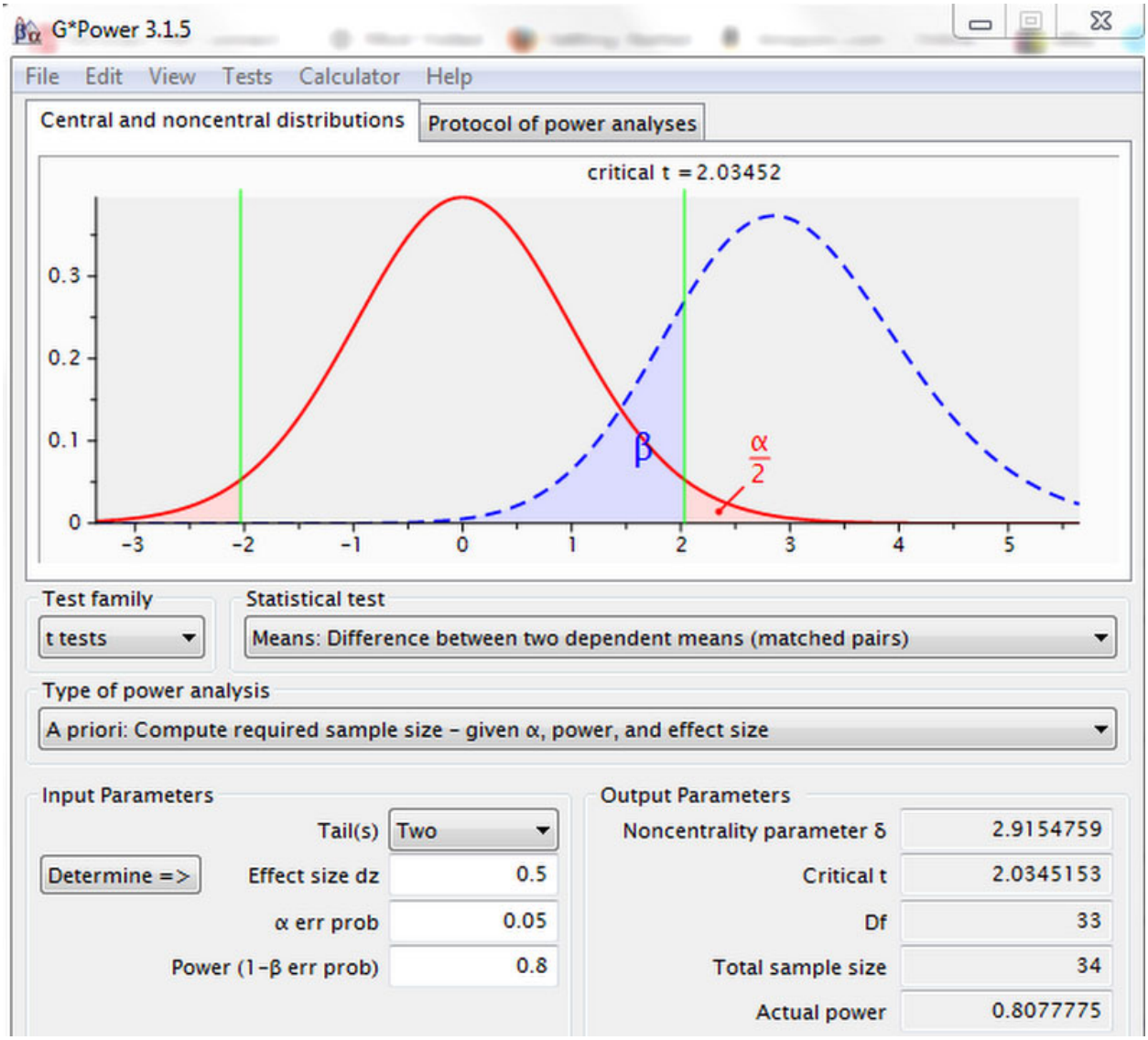

Figure 3. A priori sample size calculation of the study. *Note*: Original graphic generated in G*Power software (Faul et al., 2009) on the basis of Cohen's (2013) recommendations.

In qualitative methods, a sample size of 12-16 is sufficient to reach data saturation, that is, the point at which participants begin to repeat themes (McNabb, 2010). Therefore, a minimum

sample of 12-16 teachers was sought for the qualitative portion of the study. All teachers participating in the quantitative part of the study were offered an opportunity to participate in the qualitative part of the study. All survey respondents were asked if they were willing to sit for a follow-up interview. All teachers from the original sample were asked to participate in a follow-up interview.

Data Collection

Data collection for the study is described in two sections. First, the instruments of the study are described. Second, the actual means of gathering data on the basis of the instruments of the study is described.

Instruments

The formal instrument (Appendix A) used in the study was the ATI-R (Trigwell et al., 2005). The ATI-R consisted of 22 questions that measure different kinds of teaching practices (in particular, student- and teacher-oriented pedagogy). The scoring devise (Appendix B) separates both types of teaching practices for further analysis. Teaching practices were not content specific and therefore appropriate for use across multiple disciplines.

The reliability of the ATI-R was initially measured through Cronbach's Alpha, the standard measure (Santos, 1999) of the internal reliability of scales, and found to have an Alpha of .82. Therefore, the ATI-R had a sufficiently high internal reliability for inclusion in the study. However, exploratory factor analysis had not reached uniform conclusions about the construct validity of the ATI-R. For example, Goh et al. (2014) were able to extract five components from the ATI-R, suggesting that the ATI-R did not measure a single construct of teacher practice. In Trigwell, Prosser, and Waterhouse's (1999) initial study, the ATI-R was described as measuring two constructs, teacher- and student-oriented pedagogy. However, as Goh et al.'s (2014) results

indicate, it was not clear whether ATI-R measures these two constructs, or, indeed, a single construct of teacher practices. For this reason, the ATI-R was subjected to exploratory factor analysis.[1]

Upon completion of the survey, quantitative data were analyzed for patterns. The qualitative instrument was developed to further explore developing themes. In some instances, specific questions were asked based on individual survey responses.

Data Collection Procedures

Sampled individuals were asked to complete the ATI-R. The instrument was designed so that one set of responses elicited pre-VAM data and the other elicited current, or post-VAM data. All participants were asked to complete their instrument on the basis of both their current (post-VAM) practices but also on the basis of their practices, as best as they could be remembered, for the pre-VAM era. Each instrument had questions organized to capture pre-VAM and current (post-VAM) practices. A vignette attached to the instrument reminded individuals that, for the purposes of the pre-VAM responses, they were asked to remember their older attitudes and orientations:

> This inventory is designed to explore a dimension of the way that academics go about teaching in a specific context or subject or course. For this exercise, teacher responses should be geared towards core subject areas and how your instruction relates to the Value-Added (VAM) Growth Model being utilized at (the school), both currently and when it was implemented during the 2014-15 school year. This may mean that your

[1] A factor analysis was conducted on both teacher and student-oriented survey items. Likely due to limited power, exploratory analysis did not provide reliable constructs. For this reason, an item analysis was conducted.

responses to these items in one context may be different to the responses you might make on your teaching in other contexts or subjects.

The only demographic information collected from survey participants was the subject area (English, mathematics, social studies, science) in which they were teaching. This was done to afford respondents anonymity and to assist with an initial analysis of the quantitative data and development of questions to guide the qualitative portion of the study. Interview protocols used a semi-scripted format and lasted anywhere from 30 to 60 minutes. Interviews were conducted in the school and scheduled at the convenience of teachers. This was done before school, after school, and on plan (prep) periods, in the school office or the teacher's classroom. Interviews were recorded, transcribed verbatim, and cleaned for accuracy.

Data Analysis

Data analysis was described and justified separately for each of the research questions of the study. Finally, the overall purpose of data analysis was described as part of Dixon's (2013) knowledge generation framework. All quantitative data analysis and supporting graphics were carried out in Stata software, version 14.

RQ1 Data Analysis

The first research question of the study is as follows: What changes in teacher practices have occurred since the administration of a VAM in a particular school? The null hypothesis associated with the first research question was that teacher practices before the administration of a VAM in a particular school were not statistically different to teacher practices after the administration of a VAM in a particular school. The alternative hypothesis associated with the first research question of the study is that teacher practices before the administration of a VAM

in a particular school were statistically different to teacher practices after the administration of a VAM in a particular school.

The first research question of the study was answered through a paired-samples *t*-test in which the ATI-R score of each teacher reflecting that teacher's memory of the current-period ATI-R score. If the resulting *p* value of the comparison was below .05, the null hypothesis of the first research question of the study was rejected. Because of the structure of the ATI-R, it cannot necessarily be concluded that a higher score represented better teaching or that a lower score represented worse teaching. However, a comparison of individual scores on the 22 items of the ATI-R allowed for more precise assessments of how the VAM might have prompted changes in teaching practice. In addition, items were compared via construct. For example, one of the items on the ATI-R stated, "I structure my teaching in this subject to help students to pass the formal assessment items." Comparing the mean ATI-R response to this question before and after the VAM implementation offered insight into whether the VAM prompted teachers in the district to teach to the standardized test.

The paired-sample *t*-tests for the individual ATI-R question comparisons were not considered as part of hypothesis testing. They were utilized in order to gain more precise insight into which aspects of teaching practice, if any, changed after the implementation of the VAM.

RQ2 Data Analysis

The second research question of the study is as follows: Did VAM alter the practices of teachers? Why or why not? This research question was analyzed using explanatory mixed methods techniques to identify themes (Creswell & Plano Clark, 2011). Interview questions were developed from data that emerged from survey responses and from concepts associated with the larger VAM movement in the United States. All interviews were recorded and transcribed.

Answers to each question were analyzed and codes were developed after an answer was provided by teachers in multiple instances.

Interview Coding Scheme

Table 1

Coding Scheme

Code	Frequency	Data Examples
Students Restructuring Existing Knowledge*	15	"The one thing that comes to mind immediately is that it's part of our evaluation." (Math Teacher) "I think it comes back to what teachers describe as 'the why'." (Math Teacher) "I present information in a lot of different ways so students can reconsider their point of view." (Science Teacher)
Students Finding Their Own Learning Resources*	14	"I believe this is almost exclusive to the increased use of technology." (English Teacher) "We don't have to provide sources for them anymore due to technology." (English Teacher) "That's due to the integration of technology." (Social Studies Teacher)
Pressure to Meet Accountability Measures	24	"You're always nervous when you first open them (test results) up." (Social Studies Teacher) "I do view the scores with a sense of pride." (English Teacher) "Pressure? Absolutely. Yes." (Math Teacher)
Student Achievement in Accountability Era	36	"Oh, yeah. Definitely. My salary depends on how my kids perform on the ISTEP." (Math Teacher) "Do I teach to the test (end-of-course assessment)? You better believe I do." (English Teacher) "I make decisions on how much bang can I get for the buck." (English Teacher)
VAM Use at School	19	"Once this was identified as a priority, I tried to develop the best way to promote skill growth in my classroom." (Social Studies Teacher) "I believe they're being used (VAM results) to have a conversation, about what we believe about ourselves as a school." (Math Teacher) "I like the way that it is discussed with department meetings or in smaller PLC groups." (English Teacher)
VAM Role in Teacher Evaluation	29	"Not all of it, but certainly a fraction of it can. Absolutely. Definitely." (Social Studies Teacher) "Yes. (Long silence) It's challenging, but I think it can happen." (Science Teacher) "That's tough. To an extent, yes. I think it can be thought about as a piece of the puzzle. (Math Teacher)
Change in Teacher Practice Since VAM Implementation	22	"We started experimenting with these types of lessons when we started to consider giving every student a laptop." (English Teacher) "This (VAM) has caused me to change my instruction into a skill-based based solely on student achievement." (Social Studies Teacher) "The growth model is why we did it. If we were going to spend the time planning for the test, it was nice to see the work pay off in student scores." (Math Teacher)

Codes were then developed into themes for analysis in Chapter 4. The collected themes were synthesized into an overall exploration of why or why not the VAM altered the practices of teachers and if those practices changed, how.

Data Analysis and Knowledge Generation

Dixon (2013) described knowledge generation in terms of a sequential processing from data to information, and from information to knowledge, in the following manner: Information [is] data that is 'in formation'—that, data that has been sorted, analyzed, and displayed, and is communicated through spoken language, graphic displays, or numeric tables. Knowledge, by contrast, is defined as the meaningful links people make in their minds between information and its application in action in a specific setting (Dixon, 2013, p. 13).

In this study, data were gathered in the form of the ATI-R survey results. These data, through the data analysis processes described above, were transformed into information. In the fifth chapter of the study, the information generated through the research questions was further interpreted and compared to existing empirical results and theoretical frameworks, and transformed into knowledge. The purpose of knowledge generation was to transform the information of the study into practice guidelines for the school. If, for example, the VAM was found to not result in a meaningful change in teacher practice, then decision-makers would have a rationale to change the VAM or abandon it. On the other hand, if the VAM was found to result in a change in teacher practice, and if this change were held to be positive, then decision-makers would have an empirical basis for retaining the VAM. In either case, the information of the study is transformed into usable knowledge for decision-makers. In terms of blending the findings, a sequential explanatory approach (McNabb, 2010) was used. In this approach, the first stage was to determine, through quantitative analysis, whether the VAM actually altered

educational personnel's practices. The second stage was to understand, through qualitative analysis, why and how such changes did or did not take place.

Reliability, Validity, Trustworthiness, And Generalizability

The reliability of the ATI-R was initially measured through Cronbach's Alpha and found to have an Alpha of .82; in this study, a confirmatory Cronbach's Alpha analysis was applied to ATI-R, with a value greater than .75 considered to constitute a sufficiently high level of internal reliability, based on Santos's recommendations (Santos, 1999). An exploratory factor analysis was conducted on ATI-R in order to determine how many components can be extracted through varimax rotation (Balnaves & Caputi, 2001). Before the factor analysis, both the Kaiser-Meyer-Olkin of Measure of Sampling Adequacy (KMO) and Bartlett's Test of Sphericity was carried out in order to determine whether the sample was large enough for the factor analysis. In keeping with Balnaves and Caputi's (2001) recommendations, a minimum KMO value of 0.8 was required, and a *p* value less than .05 was required in Bartlett's Test of Sphericity. The factor analysis for ATI-R was only considered reliable if both of these conditions were met.

Both the reliability and validity of the study might be lessened by asking participants to accurately remember and report their responses as if they were completing the instrument prior to the VAM being implemented. Because the VAM was relatively recent, it was assumed that participants were able to both remember and report answers that reflect their orientations in the pre-VAM era. However, such an approach is likely to be less reliable than working from data from the pre-VAM timeframe. Because inferences based on the pre- to post-VAM comparison based on this approach were likely to be limited in strength, the validity of the study is also likely to be lower than if there were actual pre-VAM data.

The generalizability of the study might be limited by the delimitation of the study to a single site. The reliability (quantitative) and trustworthiness (qualitative) of the study might be challenged by the lack of anonymity, both real and perceived by teachers, in both the quantitative and qualitative formats. However, follow-up questions and probing were utilized to ensure that all major themes in any given interview were explored adequately.

Assumptions, Limitations, And Delimitations

The main assumption of the study was that teachers answered their instrument accurately and honestly. The study was, as explained earlier in this chapter, delimited to a single school in the United States. The study also contained a number of limitations that require further discussion.

One of the limitations of the study is the possibility of Alpha inflation (Kopalle & Lehmann, 1997) that can arise when carrying out several *t*-tests. In the context of this study, any number of paired *t*-tests can be carried out in order to compare pre-versus post-VAM impacts on each of the 22 constituent questions in ATI-R; using STATA, creating code and carrying out any number of *t*-tests is not a time-consuming task. The limitation arises not in the context of the computational difficulty of conducting several dozen individual *t*-tests, but in the likelihood of Alpha inflation. When many *t*-tests are carried out, there is a greater likelihood that a result found to be significant is not intrinsically significant, but is found significant by error (Kopalle & Lehmann, 1997). For this reason, numerous *t*-tests were carried out to examine the change in the individual ATI-R questions over time, these results, because of the possibility of Alpha inflation, were kept distinct from hypothesis testing.

An additional limitation is that of no genuine pre-intervention data. In this study, participants were asked to complete their instrument as if they were filling them out before the

VAM was implemented. Participants were therefore asked to exercise their memory, with the assumption that they could recall their attitudes and orientations from the pre-VAM era and enter them into the instrument. Such an approach is unlikely to be as accurate in measuring the effect of the VAM as an approach in which there were data actually gathered in the pre-VAM period. In the absence of such data, the only feasible manner of measuring the impact of the VAM was to ask participants to offer two sets of responses, with one set representing current responses and the other set representing the responses that they would have given had they been asked to fill out these instruments in the pre-VAM era.

Given the various limitations of a quantitative approach to the research topic, it was especially important to explain why a qualitative approach would confront similar difficulties. In the context of a subject as contentious as that of high-stakes testing and professional practice, it was unlikely that teachers would speak openly in a qualitative setting. This was especially true of this study as I also served as the principal of the school. Indeed, according to one previous finding (Busuioc & Lodge, 2016), public servants were hesitant to describe their feelings about accountability even in an anonymous setting. On the other hand, especially for teachers who were asked to complete practices questionnaires at regular intervals, it was possible that conscious aversion to accountability does not condition the given responses (especially on scales that have several items and that do not ask directly about accountability). Moreover, as noted in the first chapter, the problem of the study was a lack of knowledge about the extent to which VAMs have influenced teachers' practices, not a lack of knowledge about how teachers felt about VAMs and accountability in general. There is a consensus in the literature that most teachers are not in favor of VAMs in particular or the NCLB-based accountability structure in general (Block, 2015; Chakrabarti, 2014; Coburn et al., 2016; Franco & Seidel, 2014; Gansle et

al., 2015; Glover et al., 2016; Goldhaber et al., 2013; Grissom et al., 2014; Henry et al., 2014; Ladd, 2017; Lavery, 2016; Menken & Solorza, 2014; Paufler & Amrein-Beardsley, 2014; Plank & Condliffe, 2013; Price, 2014; Reback et al., 2014; Saultz & Saultz, 2017; Singh et al., 2015; Steinbrecher et al., 2014). There were substantial limitations to the study that would be too difficult to overcome if only using quantitative or qualitative methods. As a result, the research problem of the study necessitated both the application of a quantitative and qualitative approach.

Ethical Issues

One of the ethical issues in the research was anonymity. The data was drawn from a single school. It was reasonable to think that participants could be identified through affiliation with the school. Participants were asked to identify their gender, subject area, and years of teaching experience when completing the survey. These data were used to develop follow-up interview questions. The study did not refer to any of the participants by name.

Another ethical issue is the power dynamic that existed with me (researcher/principal) asking teachers that I was directly supervising to participate in the research. A statement was attached to the survey, included in informed consent documents, and stated at the beginning of interviews that participation in the study was voluntary and in no way would impact a teachers job status at the school.

Conclusion

The purpose of this chapter of the study was to describe and defend the research methods and design of the study. This chapter summarized a mixed-methods case study approach. Results are described in Chapter 4.

CHAPTER 4

RESULTS

The purpose of this chapter is to describe the results from both the quantitative and qualitative components of the study. First, quantitative results from the ATI-R survey are presented and analyzed. Then, themes that emerged from teacher interviews are identified, described, and analyzed. In conclusion, both components of the study are summarized as part of a transition to the fifth chapter of the study.

Purpose Statement

The purpose of this mixed-methods, case study was to examine the influence of the adoption of a single school's VAM on the practices of teachers. The presumption of the study is that the imposition of the VAM was not likely to have created changes in the practices of teachers. The purpose of the study will be achieved by answering two research questions: (1) What changes in teacher practices have occurred since the administration of a VAM in a particular school? (2) Did VAM alter the practices of teachers? Why or why not?

Research Questions And Hypotheses

The quantitative research question and accompanying hypotheses of the study are as follows:

RQ1: What changes in teacher practices have occurred since the administration of a VAM in a particular school?

H1_0: Teacher reports of their practices before the administration of a VAM in a particular school were not statistically different to teacher practices after the administration of a VAM in a particular school.

$H1_A$: Teacher reports of their practices before the administration of a VAM in a particular school were statistically different to teacher practices after the administration of a VAM in a particular school.

The qualitative research questions of the study are as follows:

RQ2: Did VAM alter the practices of teachers? Why or why not?

Quantitative Findings

The quantitative findings of the study answered the following research question: Were teacher practices before the administration of a VAM in a particular school equivalent to teacher practices after the administration of a VAM in a particular school? Teacher practices were measured on the ATI-R (Trigwell et al., 2015) instrument described in Chapter 3 and presented as Appendix A. There were 22 questions on the ATI-R, and answers to each question were analyzed through a matched-pairs *t*-test. Out of 55 teachers asked voluntarily to complete the survey instrument, 25 teachers submitted survey responses (45%). Due to not meeting the threshold of 34 respondents needed for the recommended sample size, individual factor analyses were carried for each of the respondents. There were 7 English teachers, 5 mathematics teachers, 9 science teachers, and 4 social studies that responded to the survey. To protect anonymity, age and years of experience were not documented.

Teacher Survey Findings

Table 2

Teacher Findings

ATI-R Question	Before or After Greater?	*t,* Before Minus After	Significant at $p < .05$?
1	Same	1.00	No
2	After	-1.68	No
3	Same	1.00	No
4	Before	1.01	No
5	After	-0.67	No
6	Before	0.34	No
7	**After**	**-2.02**	**Yes**
8	After	-0.67	No
9	Before	0.67	No
10	Before	0.67	No
11	Before	1.68	No
12	After	-0.67	No
13	After	-1.35	No
14	After	-0.34	No
15	After	-1.01	No
16	Before	0.34	No
17	Same	1.00	No
18	After	-1.01	No
19	Same	1.00	No
20	After	-0.34	No
21	**After**	**-2.36**	**Yes**
22	After	-0.67	No

Summary of Statistically Significant Findings

As part of the quantitative analysis, a factor analysis was conducted. The statistically significant findings from the ATI-R (Trigwell et al., 2015) survey emerged from questions 7 and 21. After VAM, teachers were more likely to encourage students to restructure their existing knowledge in terms of developing new ways of thinking about the subject, $t(23) = -2.02$, $p < .05$. Mean scores jumped from 3.8 to 4.4 on the 5 point scale used in the survey instrument. After VAM, teachers were also more likely to agree with the belief that teaching should include helping students find their own learning resources, $t(23) = -2.36$, $p < .05$. Mean

scores increased from 3.0 to 3.7 after VAM. It is possible, therefore, that the implementation of VAM positively altered teachers' conceptions regarding (a) the development of new ways of thinking and (b) the promotion of independent learning resources vis-à-vis their students.

Qualitative Teacher Findings

All 55 teachers asked to participate in the survey were contacted via email recruitment (Appendix C) to participate in follow-up interviews. 14 teachers agreed to sit for follow-up interviews. All 14 teachers were interviewed over a 3-week period at times that was convenient for them (before school, after school, prep periods, lunch hours).

Teacher Interview Characteristics

Table 3

Teacher Characteristics

Interview Date	Subject Area	Age at Beginning of 2017-18 School Year	Years of Experience at Beginning of 2017-18 School Year	Gender
1/30/19	Science	47	22	M
1/31/19	English	44	20	M
2/4/19	Social Studies	55	31	M
2/5/19	English	30	8	M
2/5/19	English	34	10	M
2/5/19	Mathematics	40	17	M
2/5/19	Mathematics	52	25	M
2/7/19	English	38	13	F
2/7/19	Social Studies	30	7	M
2/8/19	Mathematics	36	12	F
2/8/19	Social Studies	34	11	F
2/12/19	Mathematics	34	10	M
2/12/19	Social Studies	48	21	M
2/14/19	Science	27	6	F

There were 10 males (71%) and 4 females interviewed with an average age of 39.2 years old. The original sample of teachers asked to complete the survey had 47% males. The percentage of males agreeing to be interviewed is 24% higher. Further analysis would be needed to explore

any possible relationships between gender differences relative to the research questions in the study, or other factors potentially impacting a willingness to sit for a follow-up interview. The average number of teacher years of experience was 15.2 with only 3 teachers having less than 10 years of experience. All four core subject areas were represented with 4 teachers agreeing to be interviewed from each of the English, mathematics, and social studies departments. Two science teachers were interviewed. Interviews lasted between 30-50 minutes. Interview questions (Appendix D) were designed to examine the following: statistically significant findings from the quantitative part of the study, teacher perceptions on use of quantitative measures for accountability purposes, and teacher change in practice since VAM implementation. Themes that emerged from the interviews are described below.

Statistically Significant Findings from the Surveys

Students Restructuring Existing Knowledge

Survey responses indicated that after VAM implementation, teachers were more likely to encourage students to restructure their existing knowledge in terms of developing new ways of thinking about the subject and were more likely to agree that teaching should include helping students find their own learning resources. When asked to provide examples of strategies being used to encourage students to restructure existing knowledge, all teachers indicated they engage students in revision of previous knowledge by correcting errors and misconceptions while adding new information to their understanding of a topic or concept. Two other teachers indicated that this strategy was emphasized due to inclusion in the school's teacher evaluation model adopted at the beginning of the 2015-16 school year (one-year after VAM implementation) and not due to VAM. Nine teachers referenced utilizing strategies that allow students to revise knowledge that

promote student achievement and mastery learning initiatives. A science teacher summarized the rationale for this strategy particularly well:

> We can now give students multiple attempts to show mastery on a skill or topic without having to spend time grading due to technology improvements. When assigning lab reports, I will give a rubric and then I'll mark off on certain things with some comments. Students are welcome to resubmit it. I'll then grade it again. That takes time. It's valuable, but it takes time. I always want my students to have opportunities to revise knowledge. If they can learn it (skills) at some point, then we did our job. Kids don't always learn at the same rate (science teacher).

Due to teacher evaluation implications introduced within the past four years and professional development initiatives that were already underway when VAM was implemented, it is unlikely that teachers developed instructional strategies in which students were asked to revise knowledge as a result of the VAM.

Students Finding Their Own Learning Resources

Survey responses also indicated that after VAM implementation, teachers were more likely to agree with the belief that teaching should include helping students find their own learning resources. When asked, teachers provided examples of strategies being used to help students find their own learning resources. Twelve of fourteen teachers interviewed indicated that teachers likely answered the question this way due to the school launching a 1:1 technology initiative that coincided with VAM implementation. A social studies teacher summarized this point of view particularly well:

> It has to be due to technology. Students now walk into class with two computers, their phone and their laptop. They have access to all kinds of information. It seems as if it's

> now more about teaching students how to appropriately use the tools they have available to them than helping them find learning resources. Whether or not it's spoken out loud, there is an understanding that students are going to search and find things on the internet rather than go and ask their teacher.

Based on teacher feedback during interviews, it is unlikely that teacher beliefs toward helping students find their own learning resources changed as a result of the VAM.

Pressure to Meet Accountability Measures

Many teachers reported that if the success of the student was based on the results of the test, and that the success of the teacher was measured by the success of the student, therefore, teaching to the test was not only desirable, but imperative. An English teacher summarized this notion of how teachers cope with the pressure of having student achievement tied to their name:

> Pressure? Absolutely. Yes. I put everything that I have into it. I care about the kids. It can get in your head. I take everything personal. If I see a bad number or bad score, and that's my kid, I take it personally.

Student Achievement in Accountability Era

Teachers recognized the importance of their role in student achievement. A series of interview questions were asked to gauge how teachers felt about quantitative measures used to recognize student success. The questions did not mention the VAM, but referenced accountability and standardized testing. Every teacher interviewed, that taught a course in which there was a high-stakes, end-of-course assessment, made reference to planning, implementing instructional strategies, and designing assessments for the sole purpose of promoting student success on the accountability measure at the conclusion of the course. The excerpt below, from an interview with an English teacher, provides insight into the regularity of this practice:

> See, here's the deal. Is it a bad thing to teach to the test? In AP Lit, I start the year saying to students that I am 100% supportive of this test. I believe in this test. I believe in what you have to do to be successful on this test. Do I teach to the test? You better believe I do. I teach to the test, because every poem that I choose, every story that I choose, every novel that I choose, every assignment that I develop is intentional to get students to master a skill on that test.

VAM Use at School

When asked how VAM was used at the school, 12 out of 14 teachers indicated they thought results were being used for school improvement and professional development purposes. The excerpt below from an English teacher interview accentuates this belief:

> I don't think anyone is out to get me. I don't think anyone is looking at my scores and looking to fire me over them. I don't think that. I think it helps us look at the big picture. I like how we compare ourselves to schools similar to us (social studies teacher)….The intent is to try and figure out if I am performing at a level that is expected or am I performing above or below that level. I would hope that others would internalize that the same way and get some idea of whether or not we're moving as we should as a school, or as an individual teacher. This validates what we believe about ourselves as a school.

Teacher responses suggested that a professional culture has been established at the school, based on mutual respect and trust between administrators and teachers. One social studies teacher emphasized this thought during an interview session:

> Using the VAM is a direct result of trying to improve skill development. I believe in this initiative. Once administration identified this as a priority, I think a lot of us tried to develop the best way to promote skill growth in our classrooms. I have also appreciated

> that the administration has allowed us to have the latitude to express that understanding of the topic in our own way. That's the thing I like about this. It wasn't just dropped out of the sky. There was some thought that went into this. There was gradual movement and teachers were on board. This movement has worked because teachers were given many opportunities to provide feedback. We see much of what we're doing as a school in ourselves.

VAM Role in Teacher Evaluation

Teacher responses suggested that a professional culture in which data is used for continuous improvement is in place at the school. Every teacher interviewed felt VAM results could be used as part of multiple measures for evaluation purposes. When asked how this could be done, a science teacher effectively summarized the consensus view:

> I think it could inform decisions, yes. It should not be the 100% determining factor, but it could be a piece of the puzzle. Like with any standardized test, kids can have a bad day. Some kids really know the material and they're just not good test takers.

Again, it appears that teachers are amenable to include VAMs as a part of the teacher evaluation process, albeit a small one that includes multiple other measures.

Change in Teacher Practice Since VAM Implementation

Teachers were asked how their practice has changed over time, and in particular, any changes that may have occurred in the last four years (since VAM implementation). Teachers indicated that their practice did change over time, but not necessarily due to VAM. This may be due to the survey not specifically addressing areas that teachers indicated were factors in changing their practice. Both factors, identified by teachers during interviews as causes for change (technology and accountability era pressures), are relatively recent when viewed from a

historical educational perspective. One mathematics teacher summarized, particularly well, how technology was the driving force behind changes in teacher practice:

> I do think the most obvious change is due to technology. To think that we are putting our entire class into BUZZ [learning management system] is amazing. It has become a repository for all of our course resources. We started experimenting with these types of lessons 10 years ago when we started to consider giving every student a laptop. We thought the sky was falling. Everybody thought teachers were going to be replaced by machines or online courses.

Relative to technology, teachers were suspicious of change, but willing to experiment with new strategies. One social studies teacher summarized this shift in how practice changed due to experience that comes with age and repetition, in addition to improvements in technology:

> When I first got into teaching, I was like a mini-dictator. I was the single source of information that students had on a topic, or at least it seemed that way. Now, it is much more collaborative. Not only is it more common for teachers to share information with one another, but it is so easy now due to technology. Students walk into class nowadays with two computers, their laptop and their phone. Students bring-in data now that they never used to. The information at their fingertips is amazing.

Teachers admitted that attitudes changed once being exposed to, and practicing with, research-based and best practice strategies. A mathematics teacher suggested VAM was a factor in changing her practice:

> In the spring of 2015, I made that ACT binder for the Algebra II PLC. That was our end goal. That was supposedly one of the rougher classes coming through and we were expecting our data to take a hit. I made-up the binder to help teachers highlight key

> concepts to cover and hoped for the best. It was intentional. Everyone was on board and thought it was great. Our numbers didn't dip and they were even improved from the year before. The growth model is why we did it. That was encouraging. If we were going to spend the time planning for the test, it was nice to see the work pay off in student scores.

This would, once more, indicate teachers were willing to embrace new strategies that positively impact students. For most teachers, change in practice occurred when they first started teaching a course in which student success was measured by performance on an end-of-course assessment. Teacher thoughts on this topic were individually specific, yet consistent with one another. This change in teacher practice being associated with teaching a course affiliated with a high-stakes, end-of-course assessment, was particularly well explained by a social studies teacher:

> Skill over content (intentional curriculum and instructional strategy development to ensure students develop skills needed to show mastery versus emphasis on content and hoping that students develop skills along the way) is the biggest shift that I've seen in education in my career. Skill is far more valuable because it can be used over and over again. It also comes in handy on high-stakes assessments. I often times allow students to turn-in work at the end of the year from the first semester. If it's done well, I have even gone and changed a grade. It requires more grading on my end, but for kids that are willing to work, it accomplishes what I'm hoping to accomplish, skill mastery. What do I care if it takes a student a little longer to master the concepts or skills? I look better, the kid looks better, and they have learned a work ethic that they are going to take with them.

Some have argued that VAMs, or emphasis on high-stakes, standardized testing, has reduced teacher autonomy (premise for self-determination theory). While teachers may have lost some

autonomy relative to curriculum selection, when instructional practices are demonstrated as effective, and when included in the collaborative process that leads to a decision, teachers will adapt and adopt new techniques. This is consistent with the research on the factors that cause teachers to make changes to their practice, both in a practical and theoretical sense (Shugart, 2017; Springer et al., 2010). This also helps to explain how teachers have learned to adjust and thrive in this era of high-stakes testing. To that end, teachers will change and revise their teaching strategies if convinced that their students will reap the benefits.

Summary

Quantitative results of the study indicate that teacher reports of their practices before the implementation of VAM in a particular school were equivalent to teacher practices after VAM implementation. There were two changes in teacher practice identified in the survey as being statistically significant, use of strategies in which teachers asked students to revise knowledge and teachers developing lessons in which students had to identify their own learning resources. In both instances, teachers attributed these changes to a multitude of other factors. For some, the VAM may have contributed to the change, but it did not seem to be the driving force. Teachers consistently referred to school improvement and administrative initiatives that promoted skill-based instruction and mastery learning. These initiatives were developed to promote student skill development for the purpose of assisting teachers to develop strategies that would promote student achievement and growth on high-stakes, end-of-course assessments. This is logical due to the expectations placed on teachers, schools, and districts that highlight student performance on high-stakes, end-of-course assessments for accountability purposes. Additional factors that may have impacted teacher practice, along with conclusions and implications for practice, research, and theory will be discussed in Chapter 5.

CHAPTER 5

CONCLUSIONS

In Chapter 4, I presented the study findings and outlined the data analysis process, which included survey collection, analysis, and coding follow-up interviews for the purpose of establishing emergent themes. In this chapter, I discuss the major outcomes of the study along with implications for theory, practice, future research, and my personal experience and reflections on this project.

Statement of the Problem

The problem addressed in this dissertation has local as well as national components, but at both the local and national levels, the problem can be conceptualized in the same manner. For high-stakes testing to be utilized in the manner intended by NCLB and subsequent legislation, policy changes have to create an incentive for teachers to alter their practices (Coburn et al., 2016; Franco & Seidel, 2014; Gansle et al., 2015; Goldhaber et al., 2013; Henry et al., 2014; Paufler & Amrein-Beardsley, 2014; Price, 2014; Saultz & Saultz, 2017; Steinbrecher et al., 2014). If teachers do not change their practices, there is no plausible mechanism whereby they can influence student achievement. As the research literature indicates, the actual practices of educational personnel, not their emotions or other intangible factors, are associated with improved student performance (Barge & Loges, 2003; Borg, Mary, & Harriet, 2012; Davis, Gabelman, & Wingfield, 2011; Ford & Sassi, 2014; Gambrell, 2015; Griner & Stewart, 2013; Hill, Rowan, & Ball, 2005; Koh, Steers, & Terborg, 1995; Milner & Howard, 2004; Mojavezi & Tamiz, 2012; Shevalier & McKenzie, 2012; Wenglinsky, 2002; Xu, Coats, & Davidson, 2012).

In this context, the problem addressed in the study is that there is insufficient empirical knowledge about the link between VAMs and changes in the practices of teachers. In the

absence of this knowledge, legislators and policy-makers cannot be certain that statutes such as NCLB, and now ESSA, are having their intended effect of creating an accountability-based educational system in the United States. In the local context, a school that has created a VAM structure does not know whether this structure has been effective in changing the practices of teachers. The absence of local knowledge means that decision-makers lack the empirical support necessary to determine whether to keep the VAM in place, discard the VAM, or change the VAM. If, for example, the VAM has not resulted in a meaningful change in teacher practice, then local decision-makers would have a rationale to change the VAM or abandon it. On the other hand, if the VAM resulted in a change in teacher practice, and if this change were held to be positive, then decision-makers would have an empirical basis for retaining the VAM.

Purpose Statement

The purpose of this mixed-methods, case study was to examine the impact of the adoption of a single school's VAM on the practices of teachers. The presumption of the study was that the imposition of the VAM was not likely to have created changes in the practices of teachers. The purpose of the study was achieved by answering two research questions: (1) What changes in teacher practices have occurred since the administration of a VAM in a particular school? (2) Did VAM alter the practices of teachers? Why or why not?

Research Questions And Hypotheses

The quantitative research question and accompanying hypotheses of the study are as follows:

RQ1: What changes in teacher practices have occurred since the administration of a VAM in a particular school?

H1_0: Teacher reports of their practices before the administration of a VAM in a particular school were not statistically different to teacher practices after the administration of a VAM in a particular school.

H1_A: Teacher reports of their practices before the administration of a VAM in a particular school were statistically different to teacher practices after the administration of a VAM in a particular school.

The qualitative research questions of the study are as follows:

RQ2: Did VAM alter the practices of teachers? Why or why not?

Major Findings

RQ1: What changes in teacher practices have occurred since the administration of a VAM in a particular school?

The study was designed to further explore the relationship between VAM evaluation methodologies and teacher practice. Quantitative results from the study indicate that teacher reports of their practices before the implementation of VAM in a particular school were equivalent to teacher reports of their practices after VAM was implemented in a particular school. There were two questions in the quantitative part of the study in which teacher answers suggested that a change in practice had occurred. Teachers indicated their practice had changed relative to: (1) the use of strategies in which teachers asked students to revise knowledge and, (2) teachers developing lessons in which students had to identify their own learning resources.

RQ2: Did VAM alter the practices of teachers? Why or why not?

The qualitative part of the study was designed to explore whether or not teachers changed their practice after implementation of VAM. Teachers were asked specific questions during the qualitative part of the study regarding their change in practice relative to: (1) the use of

strategies in which teachers asked students to revise knowledge and, (2) teachers developing lessons in which students had to identify their own learning resources. In both instances, teachers attributed changes in practice to multiple other factors; the most reported was an increased emphasis and use of technology in recent years and school improvement and administrative initiatives that promoted skill-based instruction and mastery learning. Both initiatives were developed, in part, to promote student skill development that is highlighted by performance on high-stakes, end-of-course assessments. Teachers also indicated that the expectations, and inherent pressures, that come with teaching high-stakes courses being measured for accountability purposes have changed their practice. While VAM alone in a particular school did not seem to change teacher practice, it is plausible that school improvement and administrative initiatives developed to promote student achievement as measured by VAM and other accountability measures (within a particular school) did contribute to changes in teacher practice.

Implications for Practice

Teachers experiment with new practices in the classroom all the time. Those that are perceived to work are absorbed into the teacher's repertoire and are not often found to be a change in practice. Those that do not work are quickly discarded and not used again (Richardson & Anders, 1994; Fullan, 1993). While study results indicate VAM did not change teacher practice, technology and pressures associated with high-stakes end of course assessment did impact teacher practice. In both cases, teachers experimented with new strategies designed to increase student achievement. Change was not precipitated by the tool used to measure achievement (VAM), but due to the belief that change would benefit students.

As previously discussed in Chapter 2, the history of examining student achievement through various outputs is a long one. The Coleman Report (1966) may have introduced the concept of measuring educational outcomes to the public consciousness, but subsequent cultural and political developments in the 1980s (Sputnik and Nation at Risk), resulted in Sanders and Horn (1990s) developing a statistical model in which schools, and teachers, were expected to show year-to-year progress on academic achievement and growth. Accountability policies soon followed and by 2014, all 50 states were using growth or value-added models to measure teacher effectiveness as a component of teacher evaluation (Collins & Amrein-Beardsley, 2014). The body of research on this topic suggests that value-added measures do not change teacher practice. The quantitative component of this study confirmed this finding. Yet, responses from teachers in the qualitative part of the study indicate there is an emphasis on teaching to the test and value in the ability to thrive in a pressurized environment where performance on one-time standardized test scores is crucial for success. Some may say the two are intertwined. I disagree. Teachers are not changing practice due to value-added measures. Developing strategies that promote success on high-stakes end-of-course assessments is in the best interest of students. It just so happens that in today's educational landscape, the high-stakes exams in which many value-added models use to measure achievement and growth are the same exams that are most relevant to students. As a result, teachers prioritize student performance on these exams. Teachers have always developed strategies designed to promote student achievement. This is another example of teacher's utilizing strategies that promote student success and discarding those that do not (Richardson & Anders, 1994; Fullan, 1993).

Teachers also contend that value-added models oversimplify teacher impacts on the learning process by placing too much emphasis on one-time test scores. One area of contention

is the concept of 'burying' family background (specifically income and race) factors that influence student achievement and growth in a prior year test score as a basis for establishing a baseline for the value-added calculation. This practice is controversial at best, and rejected by many in the educational community (Amrein-Beardsley, 2008; Amrein-Beardsley & Collins, 2012; Diamond & Spillane, 2004; Marchant & Finch, 2016; Konstantopoulos & Borman, 2011; Reardon, 2011).

Critics also cite unintended consequences that can result from adopting VAM, such as negatively impacting teacher morale (Ewing, 2011) and discouraging potential teachers from entering the profession (Amrein-Beardsley, 2008; Collins, 2014; Jiang et al., 2015). This is due to VAM serving as a policing mechanism that is viewed by teachers to be punitive, rather than supportive. Weisberg, Sexton, Mulhern, and Keeling (2014) noted some VAM environments are solely punitive (that is, focused on utilizing the results of VAMs for either teacher remediation or teacher dismissal) whereas, VAM results could be utilized for positive as well as corrective purposes. The VAM environment in this study was not punitive, but implemented for school improvement and professional development purposes. It does not appear that analyses (Weisberg et al., 2014) of this kind have been carried out previously. In this case, teachers did not share in the sentiment that VAM was a policing mechanism, but a positive influence on school improvement. The implications are numerous. Most teachers in the study agreed that VAM could be part of a comprehensive evaluation model. There may be a practical application, likely at the local level, for school leaders to collaborate with teachers in developing measures that blend the best of both worlds (statistical models vs. observations and artifact collection).

This study will add to the body of research on the impact of statistical models used to measure student achievement and growth on teacher practice. While there are still many

unanswered questions about the constructs of these models, policy makers must take into consideration the criteria needed for teachers to change their practice. In short, teachers will not make changes unless they believe it is in the best interest of students. To that end, for statistical models like VAMs to have any chance of ever being accepted by educators, the assessments they are measuring must be meaningful and relevant to students.

Implications for Theory

The empirical results presented and discussed in Chapter 4 of the study have illustrated that self-determination theory (Ryan & Deci, 2000) was an appropriate theoretical framework for the study. The three innate psychological needs that self-determination theory is based on – competence, autonomy, and relatedness (Ryan & Deci, 2000) was explored through each of the components of self-determination theory in the context of the study. In terms of competence, it was possible that the adoption of VAM signaled to teachers that their competence would be measured through a single metric (one-time student test score), which might not be under the control of the teacher. To the extent that VAM reduces the various manifestations of teacher competence into a single, possible unreliable or invalid measurement, self-determination predicts that the utilization of VAMs would demotivate teachers by signaling that non-VAM-based forms of teacher competence are not important.

Insofar as VAM being perceived by teachers as an attempt to control the process and content of teaching, and thereby the freedom of the individual teacher, VAM also might reduce autonomy. As noted in Chapter 1, the drive towards teacher measurement has been closely tied to the need, as expressed in educational policy, to prepare students for higher levels of achievement on various tests, whether mandated by NCLB or state-specific tests in the era of Every Child Succeeds Act (ESSA, 2016). To the extent that preparing children for success in

high-stakes testing requires a standardization of teaching, and to the extent that the standardization of teaching reduces the freedom of the individual teacher, self-determination theory predicts that the utilization of VAMs would be harmful due to reducing teacher autonomy.

Self-determination theory can be also be applied to VAMs in terms of relatedness. The relatedness component of self-determination reflects the individual need to feel cared for and valued by others (Ryan & Deci, 2000). To the extent that VAMs might make teachers feel that they will be rigidly judged and subjected to close control, VAMs might reduce the feeling of teacher relatedness and thereby do additional damage to teachers' motivation and ability. Accountability policies that include VAM-like measures, even when designed as an incentive mechanism, are perceived by teachers as punitive. In these circumstances, teacher practice does not change and student achievement is not influenced (Barge & Loges, 2003; Borg, Mary, & Harriet, 2012; Davis, Gabelman, & Wingfield, 2011; Ford & Sassi, 2014; Gambrell, 2015; Griner & Stewart, 2013; Hill, Rowan, & Ball, 2005; Koh, Steers, & Terborg, 1995; Milner & Howard, 2004; Mojavezi & Tamiz, 2012; Shevalier & McKenzie, 2012; Wenglinsky, 2002; Xu, Coats, & Davidson, 2012). Yet, teachers at the school have learned to link their personal and professional success with the success of students. It appears the teachers have developed a cooperative learning strategy amongst themselves to deliver exemplary test results. This relatedness example provides insight into how self-determination theory is used in practice. Most teachers indicated that being able to thrive under this pressure has become part of the professional culture within the school.

Self-determination theory is rooted in motivation. As a result, the interpretative and explanatory scope of the findings have been enhanced by choosing self-determination theory to frame the study.

Implications for Future Research

VAM methodology calculates performance on standardized, longitudinal assessments in which a pre and post score can be used as benchmarks to measure student growth. In most instances, VAM was developed to inform state accountability mandates tied to teacher and school performance (Collins & Amrein-Beardsley, 2014). In this study, VAM was not developed to satisfy state accountability measures, but to inform school improvement and professional development initiatives. Unfortunately, like most VAMs developed for accountability purposes, the VAM in the study did not measure performance on an exam that was deemed most relevant to students, and in turn, teachers. VAM in the study measured student performance on longitudinal exams developed by ACT for students in grades 8-11 (EXPLORE, PLAN, ACT). While the 11th grade exam was the actual college entrance exam, prior year exams were often designated 'secondary targets' for teachers. In some instances, teachers cited state accountability exams in English and math as primary targets in which curriculum was aligned. In most cases, teachers placed emphasis on, *Advanced Placement* and dual credit courses. Future research and development of models that measure expected vs. actual growth on assessments that are most relevant to students are likely to be impactful.

VAM in this study was not developed for accountability purposes. This study may have been the first of its kind in that VAM was developed for school improvement and professional development purposes. The school in the study is considered a high-performing high school with a teaching staff that is well respected both locally and within the state of Indiana. School

and district leadership were interested in exploring different ways to measure student growth in comparison to students in like schools. This is a noteworthy distinction between the VAM in the study and those typically used for accountability purposes. The overriding goals for VAM development were: communicating student score projections (relevance for students and parents), comparing student value-added to that of like-demographic schools (measure of school effectiveness verses peers), calculating annual value-added by course and teacher; measuring effectiveness of interventions (institutional research), and comparing value-added across cohorts (measure of school improvement).

As a result, teachers did not feel as if VAM was designed for punitive measures. Teachers seemed to be more willing to accept VAM results due to the non-threatening use of scores for evaluation purposes. Future studies may be warranted to explore the relationship between VAM use for school improvement and professional development purposes.

The school in the study had a free and reduced lunch student population of 20-25% during the years in which VAM was implemented. This places the school in the top quartile of schools in the state relative to socio-economic status of students. The district passed general fund referendums in 2011 and 2018 to support programs that may have been discontinued without additional funding from the community. Teachers have a multitude of resources available to them. The challenges teachers are faced with are not due to state accountability measures or working with students and families that do not support education. As a result, teachers have flexibility in pursuing instructional strategies that may not be present in lower-income or lower-achieving schools (Ingersol, 2001). Additional research is needed to explore the relationship between teacher practice in high/low-income schools and high/low achieving schools.

Prior research conducted by Boyland, Harvey, Quick, and Choi (2014), explored professional development opportunities provided to teachers that performed low on VAM. While the intent of this study was not to examine professional development opportunities as a result of VAM scores, the development of VAM at the school in the study was for school improvement and professional development purposes. This relationship needs to be explored further. Few would argue the intent of VAM implementation is crucial to further discussions (accountability vs. school improvement) about appropriate use of VAM. Either way, there are limited professional development strategies outlined for teachers to explore that have low VAM scores. To date, VAM does not purport to show teachers ways in which their methods are flawed or how to change them. In most cases, VAM results are used to criticize and impart pressure to perform without providing a guideline or pathway towards improvement.

Researcher's Statement

As I reflect on this research and how it has impacted my professional growth, I go back to many of the conversations that took place between 2007-2010 between the high school and district leadership teams. During this time (and to some extent to this day), we were contemplating accountability era mandates and how we could leverage them for school improvement purposes. Those early conversations led to a one-of-a-kind partnership with ACT in which a growth model (VAM) was developed for the stated purpose of:

1. Comparing student value-added to that of like-demographic schools; measure of school effectiveness
2. Calculating annual value-added by course and teacher; measure of effectiveness of interventions

3. Comparing value-added across cohorts to that of like-demographic schools; measure of school improvement

The development of VAM, and subsequent initiatives that emerged in the spirit of school improvement, have driven our work for the past 12 years. I have developed a deep appreciation for the teachers and administrators who have joined me in this journey of exploring different ways to measure student, teacher, and school performance. This appreciation is built upon a commitment to pursue school improvement initiatives in the best interest of students. One teacher in particular, commented on this professional culture while being interviewed:

> I love how this school works. I don't know how my high school worked, because I only know it from a student standpoint, but I love how we approach our professional growth and our curriculum design. I think it makes sense how we use data to drive our work, but we don't use it as a weapon. We don't use data to punish or intimidate. I think it's good that we're really cognizant of what skills we want students to master and that we encourage teachers to own those skills.

The commitment from school and district leaders to foster a positive professional culture built around collaboration, mutual respect, empathy, and data-driven decision making should not be overlooked. It is likely a big reason that teachers were comfortable enough to accept a VAM for school improvement purposes. I thoroughly enjoyed the interview component of the study. The insight gained in listening to teachers for 30-45 minutes was priceless. I came away from those interviews impressed, not only with teachers on an individual basis, but on their collective impact in the school.

Summary

In this culminating chapter, I have provided an analysis of my understanding of how teacher practice has been impacted after VAM implementation at a school and implications for theory, practice and future research. Findings in the mixed-method study indicate that teacher practice did not change as a result of VAM implementation. This is consistent with prior research on the topic, even though this study was designed to evaluate changes in teacher practice after VAM implementation for school improvement purposes instead of for accountability reasons. This context is important in that it adds to the body of research, both theoretical and practical, stating teacher practice will not change in response to external, negative, factors, but only when changes are believed to be relevant and meaningful for students.

The overarching quantitative findings from the study are best summarized based on the statistically significant answers to questions from the survey responded to by 25 of 55 eligible teachers (those teaching core subjects at school during time of study). Teachers found that they were more likely to encourage students to restructure existing knowledge in terms of ways of thinking about a subject after VAM was implemented at the school. In addition, teachers were more likely to agree with the belief that teaching should include helping students find their own learning resources after VAM implementation. In both instances, changes after the introduction of VAM were explained as being unrelated to VAM during the qualitative component of the study (teacher interviews).

Qualitative findings were gleaned through a series of 14 interviews conducted with teachers (from the original pool of 55) after survey results were analyzed. During interviews, teachers indicated that strategies were developed that allows students to revise knowledge and promote achievement through mastery learning initiatives. Teachers also stated that student

outcomes, particularly on high-stakes, end-of-course assessments (and the associated pressures that come with them) impacted their practice. Teachers were aware, not only of the instructional strategies they were using, but how those strategies were used to leverage student achievement as measured on end-of-course assessments.

As I conclude this project, I contemplate the changes the educational system in America has undergone in the last century. Specifically how statistical measures, such as VAM, have taken on increased importance in the eyes of many. This study was designed to examine the association with teacher practice after VAM implantation in a particular school. To that end, I hope this study added to the body of research on the factors that truly impact teacher practice and student achievement.

References

Achinstein, B., & Ogawa, R. T. (2012). New teachers of color and culturally responsive teaching in an era of educational accountability: Caught in a double bind. *Journal of educational change, 13*(1), 1-39.

ACT. (2008). The forgotten middle. Ensuring that all students are on target for college and career readiness before high school. Iowa City, IA: Author.

Akos, P., Lambie, G., Milsom, A., & Gilbert, K. (2007). Early adolescents' aspirations and academic tracking: An exploratory investigation. *Professional School Counseling*, *11*(1), 57-64.

American Statistical Association. (2014). ASA statement on using value-added models for educational assessment. Alexandria, VA: Author.

Amrein-Beardsley, A. (2012). Recruiting Expert Teachers into High-Needs Schools: Leadership, Money, and Colleagues. *Education Policy Analysis Archives*, *20*(27).

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, *37*(2), 65-75.

Amrein-Beardsley, A., & Holloway, J. (2019). Value-added models for teacher evaluation and accountability: Commonsense assumptions. *Educational Policy*, *33*(3), 516-542.

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (EVAAS) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives, 20*(12).

Anderman, E.M., Anderman, L.H., Yough, M.S., & Gimbert, B.G. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist*, *45*(2), 123-137.

Angrist, J., Hull, P., Pathak, P., & Walters, C. (2016). Interpreting tests of school VAM validity. *American Economic Review*, *106*(5), 388-92.

Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry*, 11(1), 3-42.

Archbald, D.A., & Newmann, F.M. (1988). Beyond standardized testing. *Reston, VA: National Association of Secondary School Principals.*

Atkinson, B. M. (2015). Teachers' practices: responding to governmentality in accountability testing policy. *International Journal of Leadership in Education, 18*(1), 34-60.

Ballou, D, & Springer, M. G. (2015). Using student test scores to measure teacher performance some problems in the design and implementation of evaluation systems. *Educational Researcher*, *44*(2), 77-86.

Balnaves, M., & Caputi, P. (2001). *Introduction to quantitative research methods: An investigative approach*. Thousand Oaks, CA: Sage.

Barge, J. K., & Loges, W. E. (2003). Parent, student, and teacher perceptions of parental involvement. *Journal of Applied Communication Research, 31*(2), 140-163.

Barnett, J. H., & Amrein-Beardsley, A. (2011). Actions over credentials: Moving from highly qualified to measurably effective [Commentary]. *Teachers College Record*.

Battelle for Kids. (2011). *A guide for educational leaders*. Columbus, Ohio: Battelle for Kids.

Berger, A. A. (2013). *Media and communication research methods: An introduction to qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, *116*(1), n1.

Bernard, H. R., & Bernard, H. R. (2012). *Social research methods: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Bishop, J. H., & Mane, F. (2001). The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success. *Labour Economics*, *8*(2), 203-222.

Block, C. R. (2015). Examining a public Montessori school's response to the pressures of high-stakes accountability. *Journal of Montessori Research, 1*(1), 42-54.

Bondy, E., Ross, D. D., Hambacher, E., & Acosta, M. (2013). Becoming warm demanders: Perspectives and practices of first year teachers. *Urban Education, 48*(3), 420-450.

Bonner, E. P. (2014). Investigating practices of highly successful mathematics teachers of traditionally underserved students. *Educational Studies in Mathematics, 86*(3), 377-399. doi:10.1007/s10649-014-9533-7

Borg, J. R., Mary, O. B., & Harriet, A. S. (2012). Closing the achievement gap between high-poverty schools and low-poverty schools. *Research in Business and Economics, 5*, 1-10.

Boyland, L., Harvey, M., Quick, M., & Choi, Y. (2014). Teacher evaluation in Indiana. *International Journal of Educational Reform, 23*(4), 270-294.

Bracey, G. W. (2002). April foolishness: The 20th anniversary of a nation at risk. *Phi Delta Kappan*, *84*(8), 616-21.

Braun, H. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. *Institute of Education Sciences*, 4-6. Washington, D.C.

Cassell, C., & Symon, G. (2004). *Essential guide to qualitative methods in organizational research*. Thousand Oaks: Sage.

Chakrabarti, R. (2014). Incentives and responses under No Child Left Behind: Credible threats and the role of competition. *Journal of Public Economics, 110*, 124-146.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2012). Great teaching.*Education Next*, *12*(3).

Chenoweth, K. (2016). ESSA offers changes that can continue learning gains. *Phi Delta Kappan*, *97*(8), 38-42. doi/pdf/10.1177/0031721716647017.

Chudowsky, N., Koenig, J., & Braun, H. (Eds.). (2010). Getting value out of value-added: Report of a workshop. *National Academies Press*, 54.

Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher, 45*(4), 243-251.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge.

Coleman, J. S. (1966). *Equality of educational opportunity summary report* (Vol. 2). US Department of Health, Education, and Welfare, Office of Education.

Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system (EVAAS). *Educational Policy Analysis Archives*, *22*(98).

Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, *116*(1), 1-32.

Collinson, D., & Tourish, D. (2015). Teaching leadership critically: New directions for leadership pedagogy. *Academy of Management Learning & Education, 14*(4), 576-594.

Cooley Fruehwirth, J. (2013). Identifying peer achievement spillovers: implications for desegregation and the achievement gap. *Quantitative economics, 4*(1), 85-124. doi:10.3982/QE93

Corcoran, S. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. *Institute of Education Sciences*, 7-10. Washington, D.C.

Crawford, R. (2017). Rethinking teaching and learning pedagogy for education in the twenty-first century: blended learning in music education. *Music Education Research, 19*(2), 195-213.

Creswell, J. W. (2015). *Research methods*. Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Darling-Hammond, L. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. *Institute of Education Sciences*, 11-13. Washington, D.C.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44*(2), 132-137.

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). Pathways to New Accountability Through the Every Student Succeeds Act.

Darling-Hammond, L., & Adamson, F. (2014). *Beyond the bubble test: How performance assessments support 21st century learning.* San Francisco, CA: Jossey-Bass.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15.

Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us?. *Educational researcher*, *31*(9), 13-25.

Data Quality Campaign. (2019). *Growth data: It matters, and it's complicated*. Washington, D.C.

Davies, M. B., & Hughes, N. (2014). *Doing a successful research project: Using qualitative or quantitative methods*. New York, NY: Palgrave Macmillan.

Davis, H. A., Gabelman, M. M., & Wingfield, R. D. (2011). " She let us be smart:" Low-income African-American first-grade students' understandings of teacher closeness and influence. *The Journal of Classroom Interaction*, 4-16.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252-279.

Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *The Teachers College Record*, *106*(6), 1145-1176.

Dieterle, S., Guarino, C., Reckase, M., & Wooldridge, J. (2012). Proceedings from: Association for Education Finance and Policy Annual Conference. How do principals group and assign students to teachers? Finding evidence in administrative data and the implications for value-added, 1-48. *Association for Education Finance and Policy*. Boston, Massachusetts.

Dixon, N. M. (2013). *Common knowledge: How companies thrive by sharing what they know*. Cambridge, MA: Harvard Business Press.

Duffy, M. E. (1987). Methodological triangulation: a vehicle for merging quantitative and qualitative research methods. *Image: The Journal of Nursing Scholarship, 19*(3), 130-133.

Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1-27.

Every Student Succeeds Act (ESSA) of 2015, Pub. L. No. 114-95, § 129 Stat. 1802 (2016). Retrieved from https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf

Ewing, J. (2011). Mathematical intimidation: Driven by the data. *Notices of the AMS*, *58*(5),667-673.

Fagioli, L. P. (2014). A comparison between value-added school estimates and currently used metrics of school accountability in California. *Educational Assessment, Evaluation and Accountability*, *26*(2), 203-222.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

Ford, A. C., & Sassi, K. (2014). Authority in cross-racial teaching and learning (re) considering the transferability of warm demander approaches. *Urban Education, 49*(1), 39-74.

Franco, M. S., & Seidel, K. (2014). Evidence for the need to more closely examine school effects in value-added modeling and related accountability policies. *Education and Urban Society*, *46*(1), 30-58.

Fullan, M. (1993). *Change forces: Probing the depths of educational reform* (Vol. 10). Psychology Press.

Fullan, M. (2000). The three stories of education reform. *Phi Delta Kappan*, *81*(8), 581.

Gambrell, L. B. (2015). Getting students hooked on the reading habit. *The Reading Teacher, 69*(3), 259-263.

Gansle, K. A., Noell, G. H., Grandstaff-Beckers, G., Stringer, A., Roberts, N., & Burns, J. M. (2015). Value-added assessment of teacher preparation: Implications for special education. *Intervention in School and Clinic, 51*(2), 106-111.

Glazerman, S., Loeb, S., Goldhaber, D. D., Raudenbush, S., Staiger, D., & Whitehurst, G. J. (2010). Evaluating teachers: The important role of value-added. *Brookings Brown Center Task Group on Teacher Quality, 201*(1-13).

Glover, T. A., Reddy, L. A., Kettler, R. J., Kurz, A., & Lekwa, A. J. (2016). Improving high-stakes decisions via formative assessment, professional development, and comprehensive educator evaluation: The school system improvement project. *TEACHERS COLLEGE RECORD, 118*(14), 1-26.

Goh, P. S. C., Wong, K. T., & Hamzah, M. S. G. (2014). The approaches to teaching inventory: A preliminary validation of the Malaysian translation. *Australian Journal of Teacher Education, 39*(1), 2-11.

Goldhaber, D. & Hansen, M. (2010a). Is it just a bad class? Assessing the stability of measured Teacher performance. *CEDR Working Paper 2010-3*. Seattle, WA: University of Washington.

Goldhaber, D. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. The ability to act on differences between teachers: Empirical work fueling the debate over the use of value-added. *Institute of Education Sciences*, 16-19. Washington, D.C.

Goldhaber, D. D., Goldschmidt, P., & Tseng, F. (2013). Teacher value-added at the high-school level: Different models, different answers? *Educational Evaluation and Policy Analysis, 35*(2), 220-236.

Goldhaber, D., & Hansen, M. (2010b). Using performance on the job to inform teacher tenure decisions. *The American Economic Review*, 254.

Goldhaber, D., & Theobald, R. (2014). Do different value-added models tell us the same things? *Carnegie Knowledge Network, 4*, 1-15.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data [Special Issue]. *Educational Researcher*, *44*(2), 96-104.

Griner, A. C., & Stewart, M. L. (2013). Addressing the achievement gap and disproportionality through the use of culturally responsive teaching practices. *Urban Education, 48*(4), 585-621. doi:10.1177/0042085912456847

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers' work environments and job attitudes. *Educational Evaluation and Policy Analysis, 36*(4), 417-436.

Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2012). Can value-added measures of teacher education performance be trusted? *East Lansing, MI: Education Policy Center at Michigan State University*.

Haertel, E. H. (2013). William H. Angoff memorial lecture series: Reliability and validity of inferences about teachers based on student scores. *Educational Testing Service*.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, *26*(1), 5-28.

Hamilton, L. S., Schwartz, H. L., Stecher, B. M., & Steele, J. L. (2013). Improving accountability through expanded measures of performance. *Journal of Educational Administration*, *51*(4), 453-475.

Hampden-Thompson, G., Guzman, L., & Lippman, L. (2013). A cross-national analysis of parental involvement and student literacy. *International Journal of Comparative Sociology, 54*(3), 246-266. doi:10.1177/0020715213501183

Han, S. W. (2016). National education systems and gender gaps in STEM occupational expectations. *International Journal of Educational Development, 49*, 175-187.

Hanushek, E. A., Rivkin, S. G., Rothstein, R., & Podgursky, M. (2004). How to improve the supply of high-quality teachers. *Brookings papers on education policy*, (7), 7-44.

Harris, D. N., & Herrington, C. D. (2015). Editors' introduction: The use of teacher value-added measures in schools new evidence, unanswered questions, and future prospects. *Educational Researcher*, 44(2), 71-76.

Hartney, M. T., & Flavin, P. (2014). The political foundations of the Black–White education achievement gap. *American Politics Research, 42*(1), 3-33. doi:10.1177/1532673X13482967

Henderikus, S. (2010). Theory. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1498-1502). Thousand Oaks, CA: Sage.

Henry, G., Rose, R., & Lauen, D. (2014). Are value-added models good enough for teacher evaluations? Assessing commonly used models with simulated and actual data. *Investigaciones de Economía de la Educación, 9*, 383-405.

Hesse-Biber, S. N. (2012). *Mixed methods research: Merging theory with practice*. New York, NY: Guilford Press.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406.

Hitt, D. H., & Tucker, P. D. (2016). Systematic review of key leader practices found to influence student achievement: A unified framework. *Review of educational research, 86*(2), 531-569.

Ho, A. D., & Kane, T. J. (2013). The Reliability of Classroom Observations by School Personnel. Research Paper. MET Project. *Bill & Melinda Gates Foundation*.

Holloway-Libell, J. Amrein-Beardsley, A., & Collins, C. (2012). All hat and no cattle. *Educational Leadership*, 70(3), 65-68.

Indiana Public Law 221-1999. (1999).

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American educational research journal*, *38*(3), 499-534.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85-108.

Jackson, S. (2015). *Research methods and statistics: A critical thinking approach*. New York, NY: Cengage Learning.

Jiang, J.Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher 44*(2), 105-116.

Johnson, S. M. (2015). Will VAMs reinforce the walls of the egg-crate school? *Educational Researcher, 44*(2), 117-126.

Kane, T. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. Recent advances in understanding of teacher value-added. *Institute of Education Sciences*, 22-25. Washington, D.C.

Kane, T. J. & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates Foundation*.

Keppel, G., Saufley, W. H., & Tokunaga, H. (1992). *Introduction to design and analysis*. New York, NY: Macmillan.

Kober, N., & Usher, A. (2012). A public education primer: Basic (and sometimes surprising) facts about the US educational system. *Center on Education Policy*.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, *47*, 180-195.

Koedel, C., & Betts, J. (2010). Value added to what? How a ceiling in the testing instrument influences value-added estimation. *Education*, *5*(1), 54-81.

Koh, W. L., Steers, R. M., & Terborg, J. R. (1995). The effects of transformational leadership on teacher attitudes and student performance in Singapore. *Journal of Organizational Behavior, 16*(4), 319-333.

Konstantopoulos, S., & Borman, G. (2011). Family background and school effects on student achievement: A multilevel analysis of the Coleman data. *Teachers College Record*, *113*(1), 97-132.

Ladd, H. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. *Institute of Education Sciences*, 26-29. Washington, D.C.

Ladd, H. F. (2017). No Child Left Behind: A deeply flawed federal policy. *Journal of Policy Analysis and Management, 36*(2), 461-469.

Lareau, A. (1987). Social class differences in family-school relationships: The importance of cultural capital. *Sociology of Education*, 73-85.

Lavery, L. E. (2016). What parents still do not know about No Child Left Behind and why it matters. *Journal of Education Policy, 31*(3), 343-361.

Layton, L. (2014, May 13). Gates Foundation urges delay in using tests for teacher evaluation. *Washington Post*. Retrieved from: https://www.washingtonpost.com/local/education/good-teaching-poor-test-scores-doubt-cast-on-grading-teachers-by-student-performance/2014/05/12/96d94812-da07-11e3-bda1-9b46b2066796_story.html?utm_term=.c73a523d753f

Leary, M. R. (2011). *Introduction to behavioral research methods*. New York, NY: Pearson.

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources state NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231.

Lewis, S., & Hardy, I. (2015). Funding, reputation and targets: The discursive logics of high-stakes testing. *Cambridge Journal of Education, 45*(2), 245-264.

Lipscomb, S., Chiang, H., & Gill, B. (2012). Value-added estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot. Full report. *Mathematica Policy Research, Inc.,* 5.

Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American educational research journal*, *19*(3), 325-340.

Marchant, G.J., & Finch, W. H. (2016). Student, school, and country: The relationship of ses and inequality to achievement. *Journal of Global Research in Education and Social Science. 6*(4), 187-196.

McAninch, A. C. (2012). The logic of performance-based assessment; or, does value-added assessment make sense? *Teacher Education and Practice*, *25*(2), 184-195.

McBurney, D., & White, T. (2011). *Research methods*. New York, NY: Cengage.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education*, *4*(4), 572-606.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy* Analysis, 9(2), 171-178.

McNabb, D. E. (2010). *Research methods for political science*. Thousand Oaks, CA: Sage.

Menken, K., & Solorza, C. (2014). No child left bilingual: Accountability and the elimination of bilingual education programs in New York City schools. *Educational Policy, 28*(1), 96-125.

Milner, H. R., & Howard, T. C. (2004). Black teachers, Black students, Black communities, and Brown: Perspectives and insights from experts. *Journal of Negro Education, 73*(3), 285-297.

Misco, T. (2008). Was that a result of my teaching? A brief exploration of value-added assessment. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *82*(1), 11-14.

Mojavezi, A., & Tamiz, M. P. (2012). The impact of teacher self-efficacy on the students' motivation and achievement. *Theory and Practice in Language Studies, 2*(3), 483-491.

Morganstein, D., & Wasserstein, R. (2014). ASA statement on value-added models, statistics and public policy. *American Statistical Association*, 1(1), 108-110.

NAEP. (2017). Student achievement data. Retrieved from https://www.nationsreportcard.gov.

Parsons, E., Koedel, C., & Tan, L. (2019). Accounting for Student Disadvantage in Value-Added Models. *Journal of Educational and Behavioral Statistics*, 1076998618803889.

Pathak, P. A., Walters, C. R., Angrist, J. D., & Hull, P. D. (2017). Leveraging lotteries for school value-added: Testing and Estimation. *The Quarterly Journal of Economics*, *871*, 919.

Paufler, N. A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal, 51*(2), 328-362.

Peurach, D. J., & Marx, G. E. (2010). Leading systemic improvement: Confronting complexity in turnaround schools. *Journal of Cases in Educational Leadership, 13*(3), 26-36.

Pianta, R. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. Signal in a noisy system: Can VAM serve a purpose for evaluation and improvement of teaching? *Institute of Education Sciences*, 30-32. Washington, D.C.

Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal, 50*(5), 1152-1182.

Price, T. A. (2014). Teacher Education under Audit: value-added measures, TVAAS, EdTPA and evidence-based theory. *Citizenship, Social and Economics Education, 13*(3), 211-225.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity?: Rising Inequality, Schools, and Children's Life Chances*, 91-116.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy, 6*(3), 207-241.

Richardson, V. & Anders, P. (1994). A Theory of Change. *Teacher change and the staff development process: A case in reading instruction*, 199-216.

Riley, R. W. & Peterson, T. K. (2008). Before the 'either-or' era. *Education Week*, *28*(5), 1.

Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, *40*(3), 411-451.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Rockoff, H. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. *Institute of Education Sciences*, 33-35. Washington, D.C.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education*, *4*(4), 537-571.

Rothstein, J. (2012). Proceedings from: Learning from recent advances in measuring teacher effectiveness. *Institute of Education Sciences*, 36-37. Washington, D.C.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2014). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, *35*(5-6), 852-882.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68-78.

Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what I can do to change it.* Cambridge, MA: Perseus Books.

Sanders, W. L. (2006). Comparisons among various educational assessment value-added models. In *Power of Two—National Value-Added Conference.* Columbus, OH. Retrieved November 3, 2016, from http://www.sas.com/resources/asset/vaconferencepaper.pdf.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, *8*(3), 299-311.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, *12*(3), 247-256.

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension, 37*(2), 1-5.

Saultz, A., & Saultz, J. W. (2017). Measuring outcomes: Lessons from the world of public education. *The Annals of Family Medicine, 15*(1), 71-76.

Sawchuk, S. (2016). ESSA loosens reins on teacher evaluations, qualifications. *Education Week*, *35*(15), 14-15.

Schochet, P. Z., & Chiang, H. S. (2010). Error rates in measuring teacher and school performance based on student test score gains. NCEE 2010-4004. *National Center for Education Evaluation and Regional Assistance*.

Shapiro, A. R. (2015). Science education. *A Companion to the History of American Science*, 320-332.

Shaw, R. D. (2016). Music teacher stress in the era of accountability. *Arts Education Policy Review, 117*(2), 104-116.

Shevalier, R., & McKenzie, B. A. (2012). Culturally responsive teaching as an ethics- and care-based approach to urban education. *Urban Education, 47*(6), 1086-1105. doi:10.1177/0042085912441483

Shugart, K. (2017). The pedagogical influences of a VAM evaluation system from the perspectives of elementary school teachers in North Georgia: A phenomenological survey. Retrieved from https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=2581&context=doctoral

Singh, P., Märtsin, M., & Glasswell, K. (2015). Dilemmatic spaces: High-stakes testing and the possibilities of collaborative knowledge work to generate learning innovations. *Teachers and Teaching, 21*(4), 379-399.

Springer, M.G. et al. (2010). Teacher pay for performance. Retrieved from https://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND_RP1416.pdf

Steinbrecher, T. D., Selig, J. P., Cosbey, J., & Thorstensen, B. I. (2014). Evaluating special educator effectiveness: Addressing issues inherent to value-added modeling. *Exceptional Children, 80*(3), 323-336.

Stevenson, I. (2013). Does technology have an impact on learning? A Fuzzy Set Analysis of historical data on the role of digital repertoires in shaping the outcomes of classroom pedagogy. *Computers & Education, 69*, 148-158.

Strauss, A., & Corbin, J. M. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Thousand Oaks, CA: Sage Publications.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, *10*(2-3), 211-227.

Theron, L. C. (2013). Black students' recollections of pathways to resilience: Lessons for school psychologists. *School Psychology International, 34*(5), 527-539. doi:10.1177/0143034312472762

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, *16*(4), 409-424.

Trigwell, K., Prosser M. & Ginns, P. (2005). Phenomenographic pedagogy and a revised Approaches to Teaching Inventory, Higher Education Research and Development, 24, 349-360.

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher education*, *37*(1), 57-70.

Inventory, Higher Education Research and Development, 24, 349-360.

Trochim, W., Donnelly, J., & Arora, K. (2015). *Research methods: The essential knowledge base*. Boston, MA: Nelson Education.

U.S. Department of Education. (2009). *Race to the Top program executive summary*. Washington, DC: Author.

U.S. Department of Education. (2010, March). A blueprint for reform: *The reauthorization of the elementary and secondary education act*. Washington DC: Author.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly, 37*(1), 21-54.

Wang, A. H., Walters, A. M., & Thurn, Y. M. (2013). Identifying highly effective urban schools: Comparing two measures of school success. *International Journal of Educational Management, 27*(5), 517-540.

Weingarten, R. (2014). Teaching and Learning over Testing. *American Educator*, 1.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *education policy analysis archives, 10*(12), 1-30.

Wilson, S. (2003). *California dreaming: Reforming mathematics education*. New Haven, CT: Yale University Press.

Witte, J. F., Wolf, P. J., Cowen, J. M., Carlson, D. E., & Fleming, D. J. (2014). High-stakes choice: Achievement and accountability in the nation's oldest urban voucher program. *Educational Evaluation and Policy Analysis, 36*(4), 437-456.

Xu, J. Z., Coats, L. T., & Davidson, M. L. (2012). Making science homework work: The perspectives of exemplary African American science teachers. *TEACHERS COLLEGE RECORD, 114*(7), 1-13.

Yin, R. K. (2009). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Zikmund, W. G. (2003). *Business research methods*. Mason, OH: Southwestern.

APPENDIX A: Instrument

Teacher Survey

APPROACHES TO TEACHING INVENTORY-R

This inventory is designed to explore a dimension of the way that academics go about teaching in a specific context or subject or course. For this exercise, teacher responses should be geared towards core subject areas and **how your instruction relates to the Value-Added (VAM) Growth Model being utilized at CPHS**, both currently and when it was implemented during the 2014-15 school year. This may mean that your responses to these items in one context may be different to the responses you might make on your teaching in other contexts or subjects.

Subject Area (select one)

- English
- Math
- Science
- Social Studies

Years of Experience (select one)

- 4-15 years
- 16+ years

Age (select one)

- 27-40 years old
- 41+ years old

For each item please circle one of the numbers (1-5). The numbers stand for the following responses:

1 - this item was **only rarely or never** true for me in this subject.

2 - this item was **sometimes** true for me in this subject.

3 - this item was true for me **about half the time** in this subject.

4 - this item was **frequently** true for me in this subject.

5 - this item was **almost always or always** true for me in this subject.

Please circle a response for each item. Do not spend a long time on each: your first reaction is probably the best one.

1. In this subject students should focus their study on what I provide them.

Prior to VAM implementation in 2014-15

Only Rarely				**Almost Always**
1	2	3	4	5

Current (2017-18), or post-VAM implementation

Only Rarely				**Almost Always**
1	2	3	4	5

2. It is important that this subject should be completely described in terms of specific objectives that relate to formal assessment items.

Prior to VAM implementation in 2014-15

Only Rarely				**Almost Always**
1	2	3	4	5

Current (2017-18), or post-VAM implementation

Only Rarely				**Almost Always**
1	2	3	4	5

3. In my interactions with students in this subject I try to develop a conversation with them about the topics we are studying.

Prior to VAM implementation in 2014-15

Only Rarely				**Almost Always**
1	2	3	4	5

Current (2017-18), or post-VAM implementation

Only Rarely				**Almost Always**
1	2	3	4	5

4. It is important to present a lot of facts to students so that they know what they have to learn for this subject.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

5. I set aside some teaching time so that the students can discuss, among themselves, key concepts and ideas in this subject.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

6. In this subject I concentrate on covering the information that might be available from key texts and readings.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

7. I encourage students to restructure their existing knowledge in terms of the new way of thinking about the subject that they will develop.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

8. In teaching sessions for this subject, I deliberately provoke debate and discussion.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

9. I structure my teaching in this subject to help students to pass the formal assessment items.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

10. I think an important reason for running teaching sessions in this subject is to give students a good set of notes.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

11. In this subject, I provide the students with the information they will need to pass the formal assessments.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

12. I should know the answers to any questions that students may put to me during this subject.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

13. I make available opportunities for students in this subject to discuss their changing understanding of the subject.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

14. It is better for students in this subject to generate their own notes rather than copy mine.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

15. A lot of teaching time in this subject should be used to question students' ideas.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

16. In this subject my teaching focuses on the good presentation of information to students.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

17. I see teaching as helping students develop new ways of thinking in this subject.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

18. In teaching this subject it is important for me to monitor students' changed understanding of the subject matter.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

19. My teaching in this subject focuses on delivering what I know to the students.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

20. Teaching in this subject should help students question their own understanding of the subject matter.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

21. Teaching in this subject should include helping students find their own learning resources.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

22. I present material to enable students to build up an information base in this subject.

Prior to VAM implementation in 2014-15

Only Rarely **Almost Always**

1 2 3 4 5

Current (2017-18), or post-VAM implementation

Only Rarely **Almost Always**

1 2 3 4 5

APPENDIX B: Instrument

Teacher Survey Scoring Devise

APPROACHES TO TEACHING INVENTORY-R Scoring Devise

The Approaches to Teaching Inventory-Revised (ATI-R) has two scales:

- Information transfer/Teacher-focused scale (ITTF)
- Conceptual Change/Student-focused scale (CCSF)

ITTF Items 1, 2, 4, 6, 9, 10, 11, 12, 16, 19, 22

CCSF items 3, 5, 7, 8, 13, 14, 15, 17, 18, 20, 21

Scoring is based on the mean numeric response (1-5) for each item in the scale. We have not published norms, nor will we, as we have gone to some lengths in writing on the research behind this inventory, that responses to it are relational and are specific to the context in which they are collected. Teachers who adopt one approach in one context may not adopt the same one in a different context. Our main use of the Inventory has been as a source of data for analysis of associations within a specific context. For example the associations between approach to teaching and perceptions of leadership in departments, or relations between approach to teaching and student approaches to learning. Permission to use this Inventory is given, provided:

a) that its source is acknowledged in all publications (ATI)* (ATI-R)**

b) that users notify Keith Trigwell of their intention to use the inventory, and

c) that once data have been collected and used as intended that the raw results on the inventory items are available for the use of Michael Prosser and/or Keith Trigwell.

APPENDIX C: Instrument

Teacher Interview Recruitment

Researcher: Jack Pettit

Study: Teacher Perception of Locally Developed Value-Added Measures

Semi-Scripted Interview Protocol Recruitment Email

Teachers,

This past June, I invited you to participate in a study about teacher perception of locally developed Value-Added (VAM) measures. You were invited to participate in the study because you have been teaching a course that is being analyzed as part of the growth model that was implemented at the beginning of the 2014-15 school year. Only teachers that have been teaching courses associated with the growth model from 2014-15 through the 2017-18 school year were asked to participate in the study. The study aims to examine how the adoption of a single high school's Value-Added (VAM) growth model shaped and influenced teacher practice. The rationale for the study is based on the insufficient empirical knowledge about the link between VAMs and changes in the practices of teachers.

I have finished with the survey collection and analysis. I believe it would be beneficial to conduct follow-up interviews to gain a better understanding of how the growth model has shaped and influenced teacher practices. The interview will consist of 21 questions and take approximately 30-45 minutes to complete. You will be asked questions that are intended to provide additional insight into both your current (post-growth model) teaching practices but also on the basis of your practices, as best as you can remember, from the pre-growth model era. Questions will be tailored to account for individual teacher levels of experience and subject area

expertise. Your participation is completely voluntary and you can choose to remove yourself from the study at any time.

The completed study will be shared with research participants, school leadership, and district leadership. Currently, the absence of local knowledge means that decision-makers lack the empirical support necessary to determine whether to keep, discard or change the growth model. Your participation may help shape school improvement practices and related policies in the school.

To participate in a follow-up interview, please send me an email and indicate the day(s) and time (before school, prep time, lunch time, after school, evening) that are best for you to be interviewed. You will be asked to complete an Informed Consent Form (attached) before being interviewed. My intent is to have all interviews completed by March 1, 2019.

Feel free to contact me if you have additional questions or concerns.

Thank you!

Chip Pettit

Crown Point High School

1500 S. Main Street

Crown Point, IN 46307

Phone: (219) 669-9599

cpettit@cps.k12.in.us

APPENDIX D: Instrument

Teacher Interview Questions

Semi-Scripted Interview Protocol

Research Question (1) - What changes in teacher practices have occurred since the administration of a VAM in a particular school?

Research Question (2) - Why or why not did VAM alter the practices of teachers? If teacher practices changed, how?

All 55 teachers in the sample will be asked to participate in a semi-scripted follow-up interview to a survey that was collected in summer 2018. Questions have been developed and grouped into three themes (teacher directed, curriculum directed, student directed) that were explored in the survey instrument. There are two specific survey questions (#7 & #21) that will be explored in detail due to respondents indicating that attitudes have changed since the implementation of a VAM at a particular school. The interviews will explore teacher's own perceptions on their role in the classroom.

Script

Thank you for taking the time to speak with me, (name). As you may recall, I administered a survey to English, math, science and social studies teachers last June to gain an understanding of their perceptions of the ACT Growth model. As you know, survey items only give so much information which is why I wanted to talk with teachers further. I'd like to learn more about how the growth model may have altered your teaching practice. The

information collected from this interview will be used in my dissertation, but your name will remain confidential/anonymous. I'm looking forward to learning from your responses, but if I ask any question that you would prefer not to answer for any reason, just let me know and we'll move on to the next question. Do you have any questions for me? If not, let's get started...

1. **Tell me about your teaching career. What led you to teaching?**
    - This is a warm-up question designed to get the teacher comfortable talking with the interviewer.
2. **What do you enjoy most about teaching?**
    - This is a warm-up question designed to get the teacher comfortable talking with the interviewer.
3. **Walk me through a typical class period in your classroom. What types of instructional strategies do you use most often?**
    - Interviewer will be listening for themes present in ATI-R survey and will follow-up as needed.
4. **I'm curious as to how your instruction has evolved since you began teaching. Are there instructional strategies that you use today that you didn't use when you started teaching?**
    - Potential follow up question – What prompted you to change?
        - This question is designed to get the teacher thinking about their teaching and getting comfortable talking with the interviewer.

5. **How do you decide when to speed up or slow down when delivering instruction, both during a class period or within a unit of instruction?**

- Follow up question – **How much input are students given in your class to determine the pace or path of instruction?**
- This question is designed to get the teacher comfortable and in a reflective mood about their teaching.

6. **Are students given time to debate points of view in your class? If so, can you give me an example?**

- This question is designed to get the teacher comfortable and in a reflective mood about their teaching.

7. **Survey results indicated that after the growth model was implemented, teachers were more apt to encourage students to restructure their existing knowledge in terms of developing new ways of thinking about a subject. Can you give me an example of how you ask students to restructure their existing knowledge?**

- This question is designed to probe at question #7 in the ATI-R survey. Teachers indicated on the survey that post-VAM, they were more likely to encourage students to restructure their existing knowledge in terms of developing new ways to think about a subject.

8. **Survey results indicated that after the growth model was implemented, teachers were more likely to develop assignments in which students had to find their own learning resources. Can you provide some examples of any assignments or projects in which you have helped students identify appropriate learning resources?**

- This question is designed to probe at question #21 in the ATI-R survey. Teachers indicated on the survey that post-VAM, they were more likely to agree with the belief that teaching should include helping students find their own learning resources.

9. **Some might say, that in today's era of educational accountability, there is pressure for teachers to teach to the test. Have you ever felt that way? Please explain.**

- This question is designed to get teachers thinking about accountability and the impact that standardized testing has had on their instruction.

10. **How has the importance placed on standardized testing changed the way in which you plan, teach, or assess students?**

- This question is also designed to get teachers thinking about accountability and the impact that standardized testing has had on their instruction.

11. **I often hear teachers say that they don't have enough time to cover all the material they are required to cover. How do you make decisions on what to cover and with how much depth? If not you, who? What goes into making those decisions?**

- This question is designed to get the teacher talking about curriculum and deciding what, when, and how to teach important topics. In addition, the interviewer is looking to gain insight on the pressures (direct or indirect) that teachers feel regarding curriculum choices.

12. **What happens if you don't get to all the material that you intended to cover in a unit of instruction or throughout the course of the year?**

- This question is designed to get the teacher talking about decisions they are making relative to their curriculum and students.

13. **Do you teach any courses with a standardized end-of-course assessment? If so, can you describe how you approach preparing students for that exam? Does this differ from other courses that you teach? If so, how and/or why?**

- This question is designed to gain insight into how much consideration the teacher is giving to the end-of-course accountability assessment when making decisions in the classroom.

14. **How do you feel when student scores are reported on end-of-course assessments (student, teacher, or school-level), (pride or disappointment) if you taught a course associated with those scores? Please explain.**

- This question is designed to gauge how much or how little attention teachers are paying to one-time test scores.

15. **Let's talk a bit about the growth model that we are using here at the high school. The growth model was introduced at the beginning of the 2014-15 school year for the purpose of informing school improvement initiatives that pertain to student achievement and growth. Is that how you understand the growth model and its purpose? If not, let's clarify.**

- This question is designed to make sure that the teacher understands the purpose and intended use of the VAM.
- *No matter how the teacher answers the question, the interviewer will clarify the purpose and intended use of the VAM.*
- **The growth model assigns a value-added measure to a student's achievement based on school, course, or teacher**

- **In our case, the value-added is determined from expected vs. actual growth on an ACT longitudinal assessment given in the spring of the year with the purpose and intended use to inform school improvement and professional development initiatives**

16. **Growth models are developed to measure the contributions of a teacher, school or district on students achievement. Do you believe that the contribution of a teacher, school or district can be measured? Why or why not. If so, how?**

- This question is designed to get the teacher talking about the growth model from a conceptual standpoint and whether or not they believe there is use for a growth model as an evaluation or school improvement tool.

17. **How do you believe growth model results are being used at the school?**

- This question is designed to explore teacher thoughts on whether or not the growth model is being used for its intended purposes.

18. **Do you believe that a growth model can be used effectively to inform school improvement decisions? Why or why not?**

- This question will be asked if the teacher indicates that a growth model could be used in some capacity (see #13). The purpose is to determine whether or not the teacher distinguishes between use for evaluation and school improvement purposes.

19. **Do you believe that a growth model can be used effectively to inform the teacher evaluation process? Why or why not?**

- This question will be asked if the teacher indicates that a growth model could be used in some capacity (see #16). The purpose is to determine whether or not the teacher distinguishes between use for evaluation and school improvement purposes.

20. **Tell me about other ways in which teacher, school, or district effectiveness can be ascertained other than using quantitative measures like standardized test scores and growth models.**

- This question will be asked to determine if the teacher has considered alternative ways in which effectiveness can be measured (observations, accreditation process, peer review, etc.).

21. **Let's finish up by talking a bit more about how your practice has changed over time. You mentioned that you used to TBA, but now utilize TBA. Is this change related to the increased emphasis in recent years on accountability, testing, the growth model or other? Please explain.**

- This question is designed to get to the actual research question guiding the qualitative portion of the study.

I really appreciate your time today – that is the end of my formal questions. Do you have anything else you would like to add about how your practice has evolved in the past four years?